CROWD MANAGEMENT SYSTEM

Capstone Project Report

MID SEMESTER EVALUATION

Submitted by:

101703544 Simran Tiwari

101703568 Sukanya Wattal

101703571 Sunishka Jain

101703573 Suruchi Pundir

BE Third Year- COE

CPG No. 46

Under the Mentorship of

Dr. Anshu Parashar

Assistant Professor

Dr. Seemu Sharma

Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

July 2020

ABSTRACT

Crowd counting is used to calculate the total number of people in images or video frames.Crowd counting is an important research topic in the field of computer vision. The crowd counting methods can be divided into three categories: the direct counting method based on target detection, the indirect method based on feature regression and crowd counting based on deep learning. This article incorporates a comparative analysis between the various techniques used in the industry and using one of them for implementation.

Emotions play an important role in human behaviour especially in large congregations of people where emotional states are prompt to be contaged and amplified. The understanding of crowd behaviour in semi-confined spaces is an important part of the design of new pedestrian facilities. Conventional manual measurement techniques are not suitable for comprehensive data collection of patterns of site occupation and movement. Real-time monitoring is tedious and tiring, but safety-critical. The complications of abnormal behavior and behavior identification are very eminent problems in video processing.

DECLARATION

We hereby declare that the design principles and working prototype model of the project entitled Crowd Management System is an authentic record of our own work carried out in the Computer Science and Engineering Department, TIET, Patiala, under the guidance of Dr. Anshu Parashar and Dr. Seemu Sharma during 6th Semester 2020.

Date:

Roll No	Name	Signature
101703544	Simran Tiwari	
101703568	Sukanya Wattal	
101703571	Sunishka Jain	
101703573	Suruchi Pundir	

Counter Signed By:

Faculty Mentor: Co-Mentor(if any):

Dr. Anshu Parashar

Assistant Professor

CSED,

TIET, Patiala

Dr. Seemu Sharma

Assistant Professor

CSED,

TIET, Patiala

ACKNOWLEDGEMENT

We would like to express our thanks to our mentors Dr. Anshu Parashar and Dr. Semu Sharma. They have been of great help in our venture, and an indispensable resource of technical knowledge. They are truly amazing mentors to have.

We are also thankful to Dr. Maninder Singh, Head, Computer Science and Engineering Department, entire faculty and staff of Computer Science and Engineering Department, and also our friends who devoted their valuable time and helped us in all possible ways towards successful completion of this project. We thank all those who have contributed either directly or indirectly towards this project.

Lastly, we would also like to thank our families for their unyielding love and encouragement. They always wanted the best for us and we admire their determination and sacrifice.

Date:

Roll No	Name	Signature
101703544	Simran Tiwari	
101703568	Sukanya Wattal	
101703571	Sunishka Jain	
101703573	Suruchi Pundir	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GUI	Graphical User Interface
APP	Application

INTRODUCTION

1.1 Project Overview

Crowd analysis is an important task that helps in the management of crowds for safety and security. Automatic crowd behaviour analysis has grown very much in the past few years. For peaceful event organizations and making sure that there are minimum casualties in the places of public and religious interests crowd behaviour analysis has become an integral part. Crowd analysis is considered to be difficult because of occlusions, background clutter, low video quality, and the difference in the human body shape and structure. As crowded scenes have a large number of people with heavy occlusions, many technologies of detection, tracking, and activity recognition, do not work in crowded scenes[5]. Despite the difficulties faced in studying crowd behavior, crowd behavior analysis has become a major interest because of its vast applications. Detecting emotions of the crowd to control the situation is an emerging interest. Crowd behavior analysis can be used for detection of abnormal behavior in situations such as panic scenarios, dangerous situations, and illegal behaviors that happen in public spaces. The emotions of people in a crowd in any environment can be estimated by their social and collective behaviors by using sensor signals like a camera, which captures and tracks the motion of people in a crowd. The understanding of crowd behaviour in semi-confined spaces is an important part of the design of new pedestrian facilities. Real-time monitoring is very difficult to do and can take some time, but it is safety-critical. The complications of abnormal behavior and behavior identification are very eminent problems in video processing. Although there is a lot more progress in human behaviour analysis over the last few years, most of the algorithms of human behavior analysis analyse individual behaviour in a very simple environment which is monitored by a single camera, but now with widespread availability of cameras and for the need of public safety, video surveillance is used from individual behavior analysis to crowd behavior analysis in multi camera networks. Emotions are brain processes which evolve to regulate behaviors under different encountered situations. Emotions play an important role in human behavior, even more so in large congregations of people where emotional states are prompt to be contaged and amplified. The decision making about these different encountered situations is governed by the evaluation dimension added by emotions, which plays a vital role in it. The first step in analyzing crowd behavior is to model the crowd. There are three major approaches in modeling crowds for crowd behavior analysis which are Microscopic approach, Macroscopic approach and Hybrid approach. The wide application domain of the crowd analysis has given rise to very recent works on abnormal crowd behavior detection based on the distribution of the magnitude of optical flow (DMOF), temporal convolutional neural network pattern (TCP), and histograms of optical flow orientation and magnitude (HOFO)[12,13,14]. Emotions play an important role in human behavior, even more so in large congregations of people where emotional states are prompt to be contaged and amplified. In computer vision tracking is the most researched topic . Most of the proposed algorithms have focused on the general problem of tracking, without specifically addressing the challenges of a crowded scene[3]. Computer vision and image processing algorithms are used for tracking humans and crowd behavior analysis based on contour detection and background modeling[4]. Emotion Regulation and Contagion is an important aspect in understanding the dynamics of emotions in groups and crowds concerns the problem of whether and how these spread or amplify across individuals in a group. Focusing on the task of emotion estimation, the framework introduced earlier addressed the recognition of individuals' membership and emotions within a group setting by means of multimodal analysis of facial and body expressions.

Crowd counting, a subtask of crowd analysis, provides density distribution which is the basic information for the crowd and the core of crowd scenes. Having a crowd count often helps analyze the importance of events. In recent years working on the crowd is becoming an interesting topic for a variety of applications such as video surveillance, public safety design and traffic control. Crowd Counting is used to count or estimate the number of people that are an image. With the help of computer vision algorithms and neural network counting of people can be done. There are many challenges in crowd counting such as low resolution and camera perspective. Some methods for crowd detection include using several sources in order to estimate a count that include face detection, texture elements, low confidence head detections and frequency domain analysis to estimate counts. Crowd counting is used to calculate the total number of people in images or video frames. Crowd counting is an important research topic in the field of computer vision. The crowd counting methods can be divided into three categories: the direct counting method based on target detection, the indirect method based on feature regression and crowd counting based on deep learning. Having a crowd count often helps analyze the importance of events. There are plenty of scenarios where crowd counting algorithms are changing the way industries work like Counting the number of people attending a sporting event, Estimating how many people attended an inauguration or a march (political rallies, perhaps), Monitoring of high-traffic areas, Helping with staffing allocation and resource allotment. There are many challenges in crowd counting such as low resolution and camera perspective. The task of crowd counting is to estimate the number of people in crowd images. Current state approaches overcome these factors by using multi-scale CNN architectures, recurrent networks and late fusion of features from multi-column CNN with different receptive fields[1]. Accurately estimating crowds from images or videos has become an increasingly important application of computer vision technology for crowd control and public safety. In some situations, like public rallies or some sports events, the number or density of participating people is an essential piece of information for future event planning and space design. The methods of crowd counting can also be extended to some other domains as well, like counting bacteria or cells from images taken from microscope, estimating animals in wildlife sanctuaries or estimating the number of vehicles at traffic jams or transportation hubs.[2]. Different Computer Vision Techniques for Crowd Counting are Detection based methods, Regression based methods, Density estimation methods and CNN based methods[11]. YOLO (You only look once) is the better method of all the methods as it has superb speed, it’s incredibly fast and can process 45 frames per second. YOLO also understands generalized object representation. It deals with object detection in a different way. It takes the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. The best thing about YOLO is that it takes the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. We predict the classes and bounding boxes of the whole image at a single run of the algorithm and detect multiple objects using a single neural network. This algorithm is simple to build and can be trained directly on a complete image.

1.2 Need Analysis

In the new year eve of 2015, 35 people were killed in a massive stampede in Shanghai, China[2]. Unfortunately, since then, many more massive stampedes have taken place around the world which have claimed many more victims. Human population in recent years is growing at a very large rate, so this increase in the growth has indirectly increased the incidences that can happen in the crowd. There has been a lot of interest in public service, security, safety and computer vision for the analysis of behaviour and mobility of the crowd.

The Crowd Surveillance System is used to count the number of people in the crowded places and detect any abnormal behavior in the crowd.

Some applications of crowd surveillance system are as follows:-

(1) Safety Control - Video surveillance cameras can be used for safety purposes in public places such as shopping malls, sports stadiums and airports by having validated monitoring of crowds for behaviour analysis, congestion analysis, and anomaly detection.

(2) Disaster Management - Some Crowd gatherings such as music concerts and political rallies are prone to face the risk of disasters such as stampedes. These situations can be avoided doing early overcrowding detection.

(3) Public Areas - There are many public places where crowd levels can be high such as malls, terminals, stations which may be affected by human health.

(4) Visual Surveillance – Public places such as playgrounds are very crowded hence Visual Surveillance system helps to reduce the anomalies by detecting abnormal behavior and alarming.

To prevent casualties, automatic detection of the situations that are critical and unusual in the dense crowd is necessary. Therefore, Crowd surveillance system can help to make decisions that are appropriate for security and safety.

1.3 Research Gaps

- The conventional methods sometimes fail for the situations like in densely crowded scenes that have severe occlusions, ambiguities and are very cluttered, where undetected abnormal activities can lead to adverse situations which are terrible. So there is a need to detect, track and analyze the behavior of crowds.
- The challenge that is faced by tracking methods is heavy occlusions in densely crowd images. This difficulty arises because of the use of a single camera which can not cover up the whole scene in the right manner.
- In [19], [20], human shape models are mentioned which can also face challenges since body components are not separated or can be congested.
- This same problem can also arise with multiple camera tracking methods [21]. To solve this new approaches can be used that can capture the data from multiple cameras and merge it together in order to extend the restricted view of a single camera.
- In [22], color histograms are used to preserve steady labeling of tracked objects for managing between the dissimilar cameras that are tracking the identical object. This approach can also face problems with denser scenes because color histograms significantly alter when the objects are occluded.

1.4 Problem Definition and Scope

In Crowd analysis applying computer vision techniques is difficult because of occlusions, background clutter, low video quality, and the difference in the human body shape and structure. Also monitoring a heavily crowded region is even more challenging because of the complex self-organizing activities, social interactions, group and individual dynamics, and the size and context of the crowd scenario, some conventional methods fail sometimes in the crowded scenes of high density because they are extremely cluttered and have high occlusions. The problem with video processing is the complications of abnormal behavior and behavior identification and Real-time monitoring is tedious and tiring. One common challenge for any CNN based crowd counting and monitoring is to meet the real-time processing requirements. In crowd counting the complexity of monitoring, tracking and counting increases as the size of the crowd increases. These are the problems which are faced during crowd analysis.

1.5 Assumptions and Constraints

1.5.1 Assumptions

Table 1: Assumptions

S.No.	Assumptions
1	Human body appearance: The detection of humans in a frame is computed through the processing of the frame based on the assumption of the shape, size, symmetry and ratio of the human body.

2	It is assumed that the people at the same location experience the same environmental and psychological influences.
3	One assumption about the mobile application is that it will always be used on mobile phones that have enough performance and latest operating systems. If the phone does not fill these requirements for the application, there may be scenarios where the application does not work as intended or even at all.

1.5.2 Constraints

Table 2: Constraints

S.No.	Constraints
1	The lack of publicly available standard benchmark datasets with many individuals and frequent behavior scenarios forces the majority of algorithms to be trained on non-standard datasets recorded under controlled circumstances.
2	The shortage of publicly available realistic datasets with high density crowds and various types of behaviors causes difficulty to have a reasonable common test bench to compare and fairly evaluate the strength and efficiency of the crowd behavior analysis systems in the real world scenarios.

3	The project must be completed within 8-9 months.

1.6 Standards

- IEEE Std 1855: IEEE Std 1855 is the first IEEE standard technology developed in the area of fuzzy logic. Its main characteristic is the interoperability, a design feature that enables system designers to develop fuzzy inference engines without taking into account the hardware/software constraints imposed by the specific architecture on which the system will be deployed.
- ISO/IEC 12207:2008: ISO/IEC 12207:2008 establishes a common framework for software life cycle processes, with well-defined terminology, that can be referenced by the software industry. It contains processes, activities, and tasks that are to be applied during the acquisition of a software product or service and during the supply, development, operation, maintenance and disposal of software products. Software includes the software portion of firmware.
- Web 2.0: Web 2.0 is used to describe the second generation of the world wide web, where it moved static HTML pages to a more interactive and dynamic Page 11 of 14 web experience. Web 2.0 is focused on the ability for people to collaborate and share information online.

1.7 Approved Objectives

- To study and review the existing work related to crowd surveillance in public places.
- To propose and implement a framework for an automated image processing application for depicting the crowd count or crowd density.

- To propose and implement a framework for an automated image processing application in real environment to detect crowd behaviour.
- To test and validate the system thereby develop and improve performance efficiency.

1.8 Methodology

- Crowd will be analyzed through a camera. Then the camera feed is passed through various neural networks that detect the abnormality in crowd behaviour.
- Feature extraction is one of the important steps of preprocessing. Feature extraction deals with the crowd density which is very useful in source information. The common analysis during the preprocessing is studied.
- After extracting the features from the image the next step is object tracking. Object tracking in crowds is used to minimize the constraint such as occlusion, illumination condition, appearance, color intensity etc.
- Another important process in a crowd analysis after object tracking is behavior or event recognition. It can be characterized by regular motion patterns that are direction, speed.

1.9 Project Outcomes and Deliverables

- A machine learning based model that can help to estimate crowd density and crowd count.
- A machine learning model to detect abnormal crowd behavior.
- A mobile application that can help the authorities to monitor crowds by keeping them updated with crowd count and crowd behavior reports.

1.10 Novelty of Work

The algorithms which we have used for crowd count and crowd behavior are different and we will combine both these algorithms to know crowd count and detect crowd behavior. The algorithm for crowd count is YOLO (You only look once) and for crowd behavior is optical flow.

2.1 Literature Survey

2.1.1 Theory Associated with Problem Area

Crowd analysis has been an emerging field of interest in the modern world because of the vast application domain. Crowd analysis can help in intelligently surveilling and monitoring heavily crowded regions and therefore preventing any unfortunate accident. Crowd behavior analysis has been a hot topic of study in graphics and simulation fields where the main aim is to simulate a realistic crowd behavior. We find a lot of applications of crowd analysis and scene understanding in video surveillance.

Crowd analysis is a central topic in computer vision. There are a variety of tasks including: crowd density estimation/person counting, crowd tracking and crowd behavior recognition. One of the several challenges in crowd analysis includes that one requires both informative and robust visual features from crowd videos. Although simple optical flow, tracklets, or a combination of motion and static features have been adopted[5].

The terms of crowd also known as 'mob' or 'mob rule' can be defined as a collective characteristic such as 'an angry crowd', a peaceful crowd' or 'a panic crowd'. Crowd is made up of the independent individual's parts, whereby each of them have their own objectives and behavior pattern which differ from the expected individually from its participants. In a crowded scene, the individual is much more variable and complex and needs to do some mathematical rules of behavior that might be useful to approximate the behavior [6].

2.1.2 Existing Systems and Solutions

Firstly there are the baseline methods used to extract low-level visual features from our dataset. Then, there are the experiments regarding emotion-based crowd representation for crowd behavior understanding. Note that during the experiments the evaluation protocol during all the experiments are fixed.[7].

A novel optical flow based features for abnormal crowd behaviour detection. The proposed feature is mainly based on the angle difference computed between the optical flow vectors in the current frame and in the previous frame at each pixel location. The angle difference information is also combined with the optical flow magnitude to produce new, effective and direction invariant event features. A one-class SVM may be utilized to learn normal crowd behavior[8].

The important component attributes in a crowd consists of density, location, speed, color and etc. By using the computer vision techniques, the information can be extracted either automatically or manually. Two types of sensors are used to capture the scene process including typology sensor and topology sensor. To get more accurate information in a crowd scene, the process of extraction information should be depending on the conditions of the environment such as illumination changes, handling the occlusion, multiple input channels and number of cameras, the changes of motion and detecting different characteristics either human or object[9].

A novel crowd counting model that maps a given the crowd scene to its density. Crowd analysis is compounded by a number of factors like inter-occlusion between people due to extreme crowding, high similarity of appearance between people and background elements, and large variability of camera view points[10].

A network for Congested Scene Recognition called CSRNet provides a data-driven and deep learning method that can understand highly congested scenes and perform accurate count estimation as well as present high-quality density maps. The proposed CSRNet is composed of two major components: a convolutional neural network (CNN) as the front-end for feature extraction and a dilated CNN for the back-end, which uses dilated kernels to deliver larger reception fields and to replace pooling operations[11].

2.1.3 Research Findings for Existing Literature

Table 3: Major Research Findings

S.No.	Name	Roll No.	Paper Title	Technology	Findings	Citations
1.	Simran Tiwari	101703544	CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes	Congested Scene Recognition Network	By taking advantage of the dilated convolutional layers, CSR-Net can expand the receptive field without losing resolution	arXiv:1802.10062 v4 [cs.CV] 11 Apr 2018
2	Sukanya Wattal	101703568	Single-Image Crowd Counting via Multi-Column	Multi-Column Convolutional Neural Network	Model trained on a source domain can be easily transferred to a target domain by	Shanghaitech University, 12 December

			Convolutional Neural Network		fine-tuning only the last few layers of the trained model, which demonstrates good generalizability of the proposed model.	2016
3	Sunishka Jain	101703571	Abnormal Crowd Behavior Detection using Social Force Model	Optical Flow with Interactive Forces	Social force approach out-performs similar approaches based on pure optical flow.	978-1-4244-3991 -1/09/$25.00 ©2009 IEEE
4	Suruchi Pundir	101703573	Abnormal Crowd Behavior Detection Using Novel Optical Flow-Based Features	Optical Flow using one-class SVM	The proposed method is very effective and performs competitive results compared to the state-of-the-art methods for anomaly detection.	978-1-5386-2939 -0/17/$31.00 ©2017 IEEE

2.1.4 Problem Identified

Several methods and techniques have been proposed by researchers to understand crowd behaviors for developing a safe and secure environment in order to avoid crowd congestion, public riots or terror attacks. In crowded scenes, standard computer vision techniques are not applicable in first hand manner due to high occlusion and complex background scenarios. Many computer vision algorithms exist for tracking, detecting and analyzing behavior of crowded scenes. Although they provide a good result in a low to medium density of population, it is still a challenge to deal with a high density crowd.

Also, in the analysis and the understanding of the video scenes, the main focus is on the detection of the object, its tracking and its behavior recognition, and therefore, the conventional methods fail sometimes in the crowded scenes of high density because they are extremely cluttered and have high occlusions.

Another problem was presented where they identified some issues concerning the unexpected behaviors of the crowd, and thus causing difficulty in controlling this crowd.

These points are summed up to clarify that:

- ➔ The behavior of people in a crowd is unexpected and unpredictable even if they are all involved in the same act or are gathered for the same reason, and this makes the crowd management more complicated.

- ➔ For pedestrians, the movement is so easy and flexible in comparison to vehicles and driving since they are governed by driving regulations, however, pedestrians can easily change their movements and directions whenever they want and wherever they want.

➔ Human behaviors are different, and they even change according to the situation, for instance, people‟s behavior in a low crowd will not be the same as in a medium or large crowd. The datasets that represent the crowd scenarios are deficient.

2.1.5 Survey of Tools and Technologies Used

The first step in analyzing crowd behavior is to model the crowd. There are three major approaches in modeling crowds for crowd behavior analysis:

- Microscopic approach: In the microscopic or object-based approach, crowd behavior is learned through the detection of individuals to analyze group behavior. Detection of objects, tracking trajectories, and recognizing activities in dense crowds make this approach significantly complex. The best-known example of this approach is the Social Force Model[18].
- Macroscopic approach: In the macroscopic or holistic approach, the crowd is seen as a global entity. It is primarily used for studying goal-oriented crowds. Scene modeling techniques are used for modeling crowd behavior instead of tracking individual entities. This top-down approach can be used for studying medium to high density crowds.
- Hybrid approach: In this approach, methods are inherited from both holistic as well as object-based approaches. This approach can be used for tracking humans in a very dense crowd.

The wide application domain of the crowd analysis has given rise to very recent works on abnormal crowd behavior detection based on the distribution of the magnitude of optical flow (DMOF), temporal convolutional neural network pattern (TCP), and histograms of optical flow orientation and magnitude

(HOFO)[12,13,14]. Recently, a few research papers have also been published for deep learning based crowd behavior analysis[16,17].

In this research, the focus is on determining crowd count and detecting any abnormal activity in surveillance videos/photos. Crowd density is determined by using the YOLO algorithm[15]. Any abnormal event is detected using the optical flow method.

In the social force model or the optical flow model, a grid of particles is placed over the image plane, and they are advected with the spacetime average of optical flow[18]. The interaction force, between particles, is calculated using the social force model. This force is then mapped into the image region to obtain Force Flow for every pixel in every frame. The normal crowd behavior is modeled using the Force Flow frames. Finally, the bag of words approach is used to classify frames as normal and abnormal.

2.2 Standards

- IEEE Std 1855: IEEE Std 1855 is the first IEEE standard technology developed in the area of fuzzy logic. Its main characteristic is the interoperability, a design feature that enables system designers to develop fuzzy inference engines without taking into account the hardware/software constraints imposed by the specific architecture on which the system will be deployed.
- ISO/IEC 12207:2008: ISO/IEC 12207:2008 establishes a common framework for software life cycle processes, with well-defined terminology, that can be referenced by the software industry. It contains processes, activities, and tasks that are to be applied during the acquisition of a software product or service and during the supply, development, operation, maintenance and disposal of software products. Software includes the software portion of firmware.
- Web 2.0: Web 2.0 is used to describe the second generation of the world wide web, where it moved static HTML pages to a more interactive and dynamic web

experience. Web 2.0 is focused on the ability for people to collaborate and share information online.

2.3 Software Requirement Specification

2.3.1 Introduction

2.3.1.1 Purpose

The purpose of this SRS document is to provide a detailed overview of our software product, its parameters, and goals. This document describes the project's scalability, its user interface, hardware and software requirements. It defines how a voter, authorization, developer and the audience see the product and its functionality.

2.3.1.2 Intended Audience and Reading Suggestions

This document is intended for finding a solution to crowd management in events etc. and determining any abnormality that may lead to serious consequences. Readers interested in a brief overview of the product should focus on Chapter 1 (Introduction) and Part 2.3.2 (Overall Description), which provides a brief overview of each aspect of the project as a whole. Readers who wish to explore our Crowd Management System in more detail should read on to Part 2.1 (Literature Survey), which expands upon the information laid out in the main overview. Part 2.3.3 (External Interface Requirements) and Chapter 4 (Design Specifications) offers further technical details, including information on the user interface as well as the hardware and software platforms on which the application will run and system architecture. Readers interested in the non-technical aspects of the project, which covers performance, safety, security and various other attributes can refer to Part 2.2.4 (Other Non-Functional Requirements) for clarity. For those who want to understand how we proceeded with this project can refer to Chapter 3 (Methodology Adopted). Those looking for in depth study of the project can finally refer to Chapter 5 (Conclusion and Future Scope) which discusses the future possibilities and benefits of crowd management system.

2.3.1.3 Project Scope

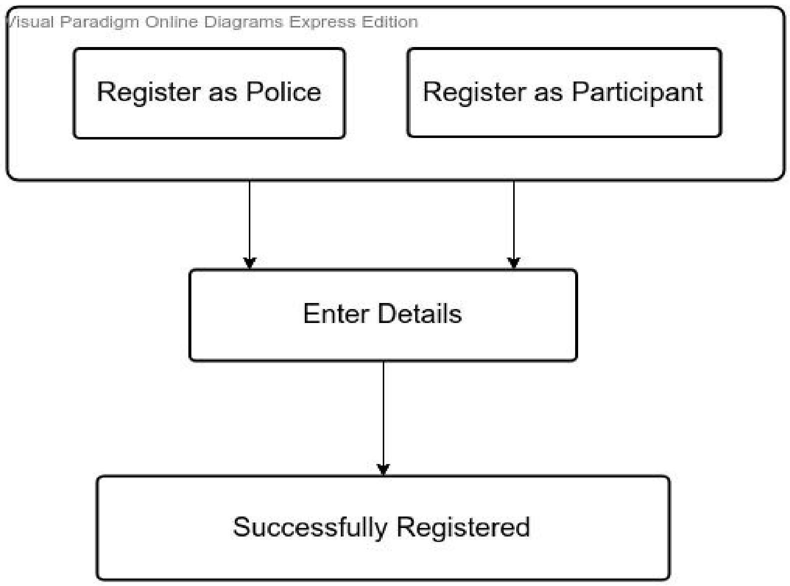
The goal is to design a hardware and a software system for Crowd Management System providing alerts for various events that the user is at and reporting the authorities if any abnormality is detected. Here, we have used deep learning and computer vision techniques for training the model and created a web application that will call the database when a user registers(or logins) into an event location and will call the script for determining crowd analysis. In case an abnormality is detected, the local authorities shall be alerted. The hardware part shall comprise of a CCTV which shall capture video frames for analysis.

The search was limited to publications in the English language addressing any of the aforementioned aspects. The aim of this contribution is to highlight advances, limitations and trends in addressing the estimation of emotions in crowds.

2.3.2 Overall Description

2.3.2.1 Product Perspective

On opening the application, the user will be asked to register(if not registered) and/or login. The interface will then ask the user to choose an event whose analysis is desired by the user. The user here can either be an authority personnel like Police, the Event Organizers etc or the public at large also referred to as Participant. The user will then be asked to select the crowd behavior that needs to be analyzed. One of the features of crowd analysis is going to be crowd count or the crowd density analysis which will help in determining whether the event area is dangerously crowded or not. If there is a threat to life and property, the authorities shall be alerted for immediate action. Also, if a participant feels that the place is too crowded, he/she can avoid visiting that area.



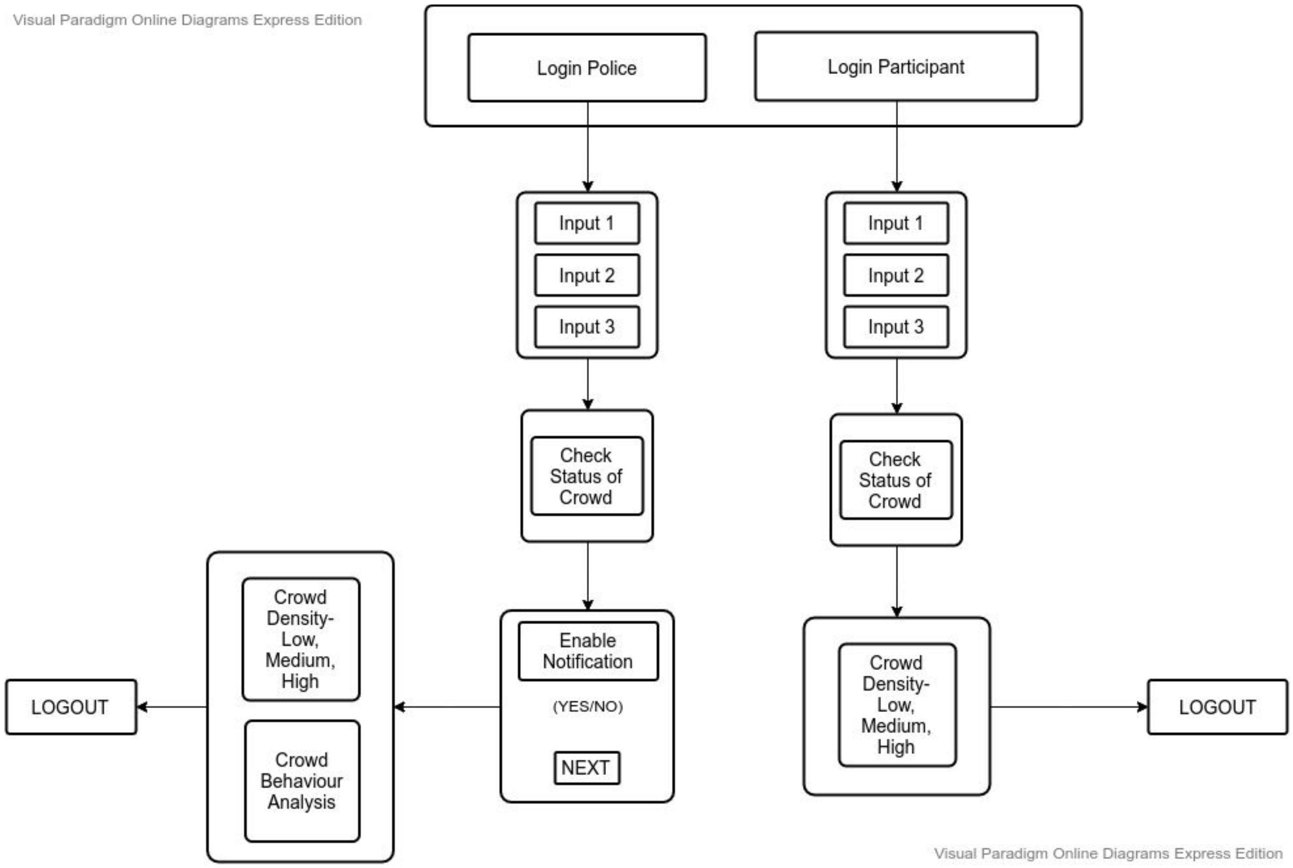


Figure 1: Product Perspective

2.3.2.2 Product Features

The following list offers a brief outline and description of the main features and functionalities of the system. Below are some of the core features which are essential to the application's operation:

- *Register*: This feature will enable the user to register either as an Authority or as a Participant. This is important to be able to enable queries from the database.
- *Login*: Once the user has registered, he/she can login to further access the feature of the app.
- *Choose Event*: To obtain the crowd analysis, the user will have to choose the event.

- *Analyse crowd*: For the user's convenience, he/she can check the crowd status so they don't get stuck in any undesirable situation w.r.t. the huge crowd. Also, an alert will be enabled if the model detects any abnormality.

2.3.3 External Interface Requirements

2.3.3.1 User Interfaces

- User friendly Graphical User Interface.
- Users should be able to accomplish every task and entry with the fewest possible keystrokes.
- The interface should enable all interaction techniques and input to be discoverable and chosen from a browse-able, hierarchical structure, arranged in order of the functions the user needs to perform.

2.3.3.2 Hardware Interfaces

- Camera should have enough resolution to capture crowd count and behavior from a distance.
- The application only runs on hardware which supports android OS.

2.3.3.3 Software Interfaces

- Android version > 4 is required to run the android application.
- The application should be able to display real-time data.

2.3.4 Other Non-functional Requirements

2.3.4.1 Performance Requirements

- Memory Consumption, Processor Consumption, Data Consumption, Battery Consumption to be managed. Commonly used resources to be used and lesser used resources to be loaded in the background. App will create minimum TCP connections or use persistent HTTP connections to reuse a TCP connection for multiple requests for proper resource management.

- The app will load in 2 to 3 seconds, rest it may vary a bit depending on processor
- Least time required to render the screen or better mobile rendering of the app. If the app functions fast but take longer to render content on the screen, it is a negative performance signal, so it is taken care of by avoiding certain factors like:

  a) Improper screen dimensions

  b) Unscaled heavy images

  c) Inconsistent font sizes

  d) Non-adaptive to different OS

  e) Excessive blocking scripts

2.3.4.2 Safety Requirements

- The system should be able to recover itself from previous crashes.
- To prevent data loss in case of system failures, the analysis should be stored in a database for a certain period of time.

2.3.4.3 Security Requirements

- ·Users must change the initially assigned login password immediately after the first successful login. Moreover, the initial should never be reused.

- ·App will be secure (free of viruses), reliable, high-performance, maintainable and scalable to such types of scenarios.

2.4 Cost Analysis

The majority of our expenses would be in setting up the hardware required for surveillance cameras. The detailed analysis is done in the table below.

Table 4: Implementation cost per unit

Item	Quantity	Approximate Cost per unit (in INR)	Total Cost (Approximate cost per unit * Quantity)
Raspberry Pi	1	3000	3100
Raspberry Pi Camera Module	1	500	500
SD card module	1	600	600
Total Cost			₹4200

Cost of implementing one hardware unit would be approximately ₹4200.

2.5 Risk Analysis

Table 5: Risk Analysis

Risk	Probability	Impact
Power Failure	Low	High
Low quality camera	Medium	Medium

No alarm sent to authorities	Low	High
False alarms sent to authorities	Low	Medium

3.1 Investigative Techniques

I. Methodologies for Crowd Count/Density Analysis

- Pixel-based Methods

These methods are based on local features for crowd count estimation in a particular image. These features may include edge detection through background subtraction. Pixel-based methods are used mostly for density estimation of the crowd rather than exact count of people due to the use of low level features. Texture is one of the important features used in identifying objects or regions of interest in an image. This high level feature is usually used in the direction of estimating the count of people as it is difficult to identify individuals in an image. In order to obtain better density estimation of the crowd in an image multi-resolution image cells are generated using a perspective projection model. Since the texture feature measurements have instabilities, the algorithm needs a uniform representation of texture features in all the divisions and thus normalization of cell size is required. An extrema searching technique in the Harris–Laplacian space is also applied. The uniformity of texture feature vectors now reduce the instability of their measurements. Each image cell division is taken as an input for feature extraction which is followed by the application of a support vector machine (SVM) to estimate the density of the crowd.

- Detection-based Methods

It takes a lot of time and training data for a machine to identify these objects. But with the recent advances in hardware and deep learning, this computer vision field has become a whole lot easier and more intuitive. Detection-based approaches use a moving-window-like detector to detect people and count their number.This method utilises head, face or human detectors, and/or segmentation algorithms to obtain the approximate location of each individual within the scene. Crowd counting is then performed directly as a subsequent step. It does Monolithic Detection, Part-based detection and Shape matching. The objective of object detection is to identify what all objects are present in the image and where they're located and second is to filter out the object of attention. The methods used for detection require well trained classifiers that can extract low-level features. Although these methods work well for detecting faces, they do not perform well on crowded images as most of the target objects are not clearly visible. To tackle this problem, researchers detect particular body parts instead of the whole body to complete crowd scenes analysis.

- Regression-based Methods

In Regression-based methods relations among extracted features from cropped image patches are studied, and then calculate the number of particular objects. Here, neither segmentation nor tracking of individuals is involved. One of the earliest attempts involves extracting the low-level features such as edge details, foreground pixels, and then applying regression modelling to it by mapping the features and the count. Localisation based methods divide an image into a number of subregions and then apply regression-based counting techniques locally. Regression-based methods only count the number of pedestrians, while ignoring the important spatial information. When executing the regression-based solution, one critical feature, called saliency, is overlooked which causes inaccurate results in local regions.

- Density estimation-based methods

A majority of the previous approaches ignored the spatial information persisting in the images. This approach incorporates spatial information in the process where the density is estimated by learning the mapping between local features and object density maps. It does linear mapping between features in the local region and its object density maps.

- CNN-based Methods

Convolutional Neural Network is a powerful neural network that uses filters to extract features from images. It also does so in such a way that position information of pixels is retained.Putting traditional approaches aside, presently, Convolutional Neural Network(CNN) based computer vision techniques are being used to achieve a better accuracy over the conventional techniques. In this technique, we build an end-to-end regression method using CNNs rather than looking at the patches of an image. Here, unlike the other methods, the entire image is taken as input in order to generate the crowd count. CNNs work really well with regression or classification tasks, and they have also proved their worth in generating density maps. The problem with using this approach is that the objects in the image can have different aspect ratios and spatial locations. In some cases the object might be covering most of the image, while in others the object might only be covering a small percentage of the image. As a result of these factors, we would require a very large number of regions resulting in a huge amount of computational time.

The other convolutional neural network are:-

1. Region-based Method (R-CNN)

Instead of working on a massive number of regions, R-CNN creates bounding boxes. RCNN uses selective search to extract these boxes from an image. At a higher level selective search looks at the image through windows of different sizes, and for each size tries to group together adjacent pixels by texture, color or intensity to identify objects. R-CNN is a very slow method as it takes time to extract so many regions for each image based on selective search and extracting features for every image region. To overcome this problem Fast R-CNN, Faster R-CNN methods came up.

2. You Only Look Once Method (YOLO)

It takes the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. We predict the classes and bounding boxes of the whole image at a single run of the algorithm and detect multiple objects using a single neural network. YOLO algorithm is fast as compared to other classification algorithms and can process 45 frames per second. YOLO algorithm is used for predicting the accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and class probabilities. The algorithm is simple to build and can be trained directly on a complete image.

II. Methodologies for Crowd Behaviour Analysis

- Crowd Motion Modelling and Segmentation

In order to make the environment more secure and safe, manage the flow of people, and also include business intelligence and marketing, analysis of crowd behaviour is necessary. Appropriate visual features need to be extracted in order to understand the crowd behaviour and distinguish regions with motion changes or similarities. Segmentation was the area to be focussed for crowd analysis in the early days. Analyzing the optical flow and modeling the movement of particles to imitate human behaviour came to be known later in other works. Crowd anomalies can be detected after learning the parameters of the crowd. Identifying the people individually can facilitate estimation of motion with a feature-based tracker or by the usage of social interaction models.

- Crowd Analysis using Deep Learning

It is difficult to perform detection, tracking, and activity recognition methods on crowded images because they have a large number of people located very close to each other with heavy occlusion and a lot of frequent movement. These existing technologies only work well with sparse crowds. The application of some existing works is limited by their scene-specific nature in which models can only make predictions for those images on which they have been trained. If the model needs to be applied on a different image, it must be trained on that particular image first. Considering the field of computer vision, some generic crowd properties, like uniformity, density of crowd, stability and collectiveness have been proposed by researchers. These properties aid the scene-independent analysis of crowd in which a generic model trained on a particular image can be applied to any other image with a crowd of people. In recent studies, the locations, events and subjects of the crowd were characterized by a more comprehensive set of 94 attributes. Keeping in mind the challenges proposed by the field of computer vision, deep learning models have succeeded largely in recent years. It is due to the availability of large scale training data that the deep learning models have become a huge success. There are no datasets available which are suitable for training generic neural network models applicable to different scenes due to their limited size, diversity of scene, etc. The learning of dynamic feature representations from video datasets in order to perform crowd analysis is still challenging for convolutional networks. The process of training models on videos is much more computationally complex than training models on images. Crowd analysis still requires new architectures and training strategies for better performance.

- Object(Person) Detection and Tracking in crowded scenes

In this method, every person in the crowd is tracked individually in order to analyze the crowd behaviour. Keeping a track on every person separately creates issues like highly increased time of computation when the number of people in the crowd becomes large. Also, the larger the crowd, the more would be the occlusions and the less would be the performance efficiency. The bottom-up approaches of separate detection and tracking of human body parts can be applied in order to reduce the problems created by occlusion. Crowd density constraints should be imposed so that the performance of detection of people of the model can be improved.

- Texture-based Approach

Researchers proposed an approach for abnormal behavior detection in crowds which could learn a shape model in order to estimate the accurate number of people in the scene. Studies showed that images of dense crowds have fine (low-level) textures and images of sparse crowds have coarse texture. For extracting crowd density features, Gray Level Dependence Matrices (GLDMs) was applied on the acquired images in a research work. This statistics based method used four measures namely energy, homogeneity entropy and contrast. This different approach towards abnormal behaviour detection, though working, was not effective enough to be chosen over earlier proposed models.

3.2 Proposed solution

In this project, we integrate both the crowd count and behaviour analysis in order to complete the crowd management system which was not implemented before in any other study. You only look once (YOLO) is a CNN based algorithm used to detect objects in an image or video file. We use this algorithm to detect the person class in the input image or frames which takes the whole image as an input in a single run. It takes the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. We predict the classes and bounding boxes of the whole image at a single run of the algorithm and detect multiple objects using a single neural network. YOLO algorithm is fast as compared to other classification algorithms and can process 45 frames per second. YOLO algorithm is used for predicting the accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and class probabilities. The algorithm is simple to build and can be trained directly on a complete image. Crowd behaviour analysis is accomplished by using an optical flow algorithm, A mobile application has been built which integrates both the crowd count and behaviour analysis and shows the required outputs to the users. The two users of the application, the participant and the security personnel, will be able to register themselves according to the event and then access the details of the crowd for successful crowd management.

3.3 Work Breakdown Structure

This project has various modules with major ones being Identification of topic and Determination of work flow, Study of neural networks and other technologies, implementation of crowd count algorithm, integration and mobile app, implementation of crowd analytics, design optimisation and testing and performing modifications. Out of these the first four have been implemented and the work has gone well according to the breakdown structure.

Table 6: Work Breakdown Structure for Sixth Semester

Activity	Month	February		March				April				May				June				July			
	Week no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Identification of topic and Determination of work flow	Plan																						
	Actual																						
Study of neural networks and other technologies	Plan																						
	Actual																						
implementation of crowd count algorithm	Plan																						
	Actual																						
integration and mobile app	Plan																						
	Actual																						

Table 7: Work Breakdown Structure for Seventh Semester

Sr. No.	Activity	Month	August				September				October			
		Week no.	23	24	25	26	27	28	29	30	31	32	33	34
5	implementation of crowd analytics	Plan												
		Actual												
6	design optimisation	Plan												
		Actual												
7	testing and performing modifications	Plan												
		Actual												

3.4 Tools and Technology

Following is the tentative list of tools and technology to be used while making the Crowd Management System. Some of the tools could be modified based on outcomes. As of now, given below is the list.

- Numpy
  - Working with arrays
- matplotlib.pyplot
  - Plotting graphs
- OpenCV
  - CV (Computer Vision) library
- Easygui, Tkinter
  - For creating GUI
- OS
  - For manipulating files
- Pillow
  - Image manipulation
- String
  - For punctuations
- Itertools
  - For count
- YOLO
  - For object detection
- Anaconda Navigator & Python IDE
  - Python offline coding
- Google Colab
  - Cloud coding
- Neural Networks (Keras, Tensorflow, CNN)
  - For classification and training

4.1 System Architecture

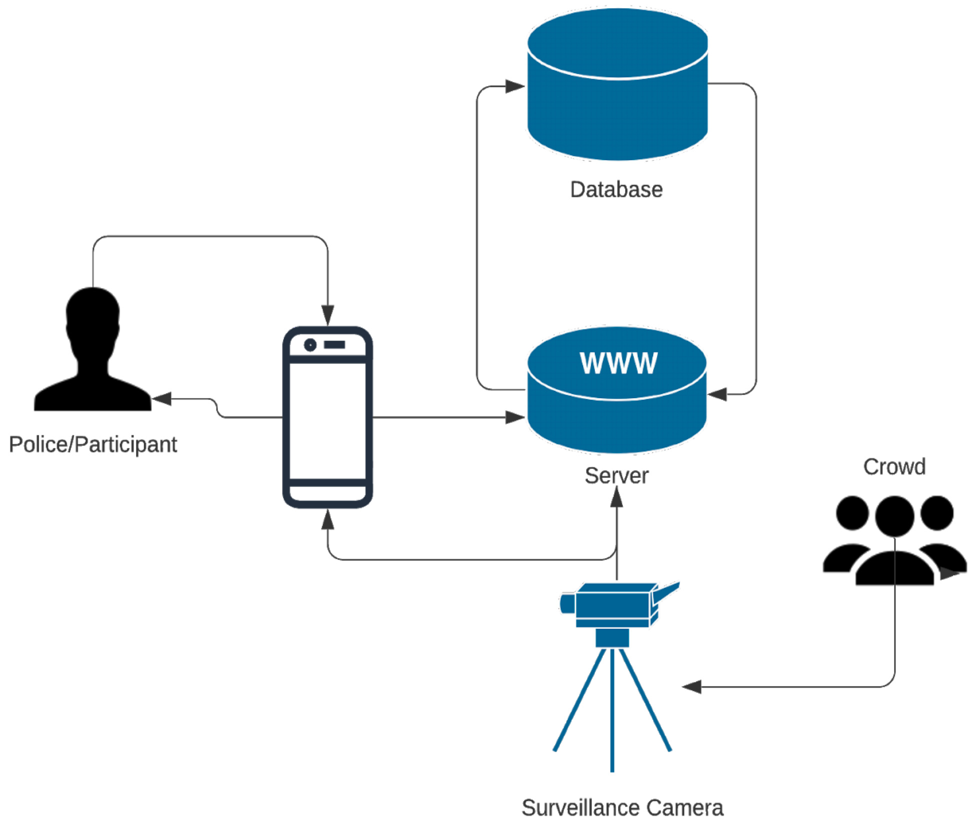


Figure 2: System Architecture Diagram

4.2 Design Level Diagrams

4.2.1 Data Design Diagram

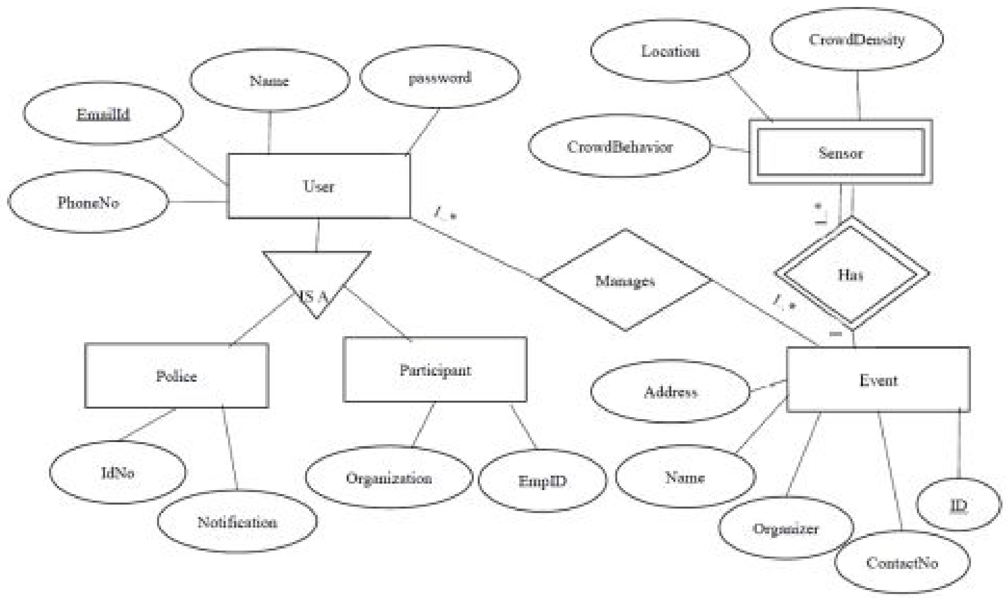


Figure 3: Data Design Diagram

4.2.2 Use Case Diagram

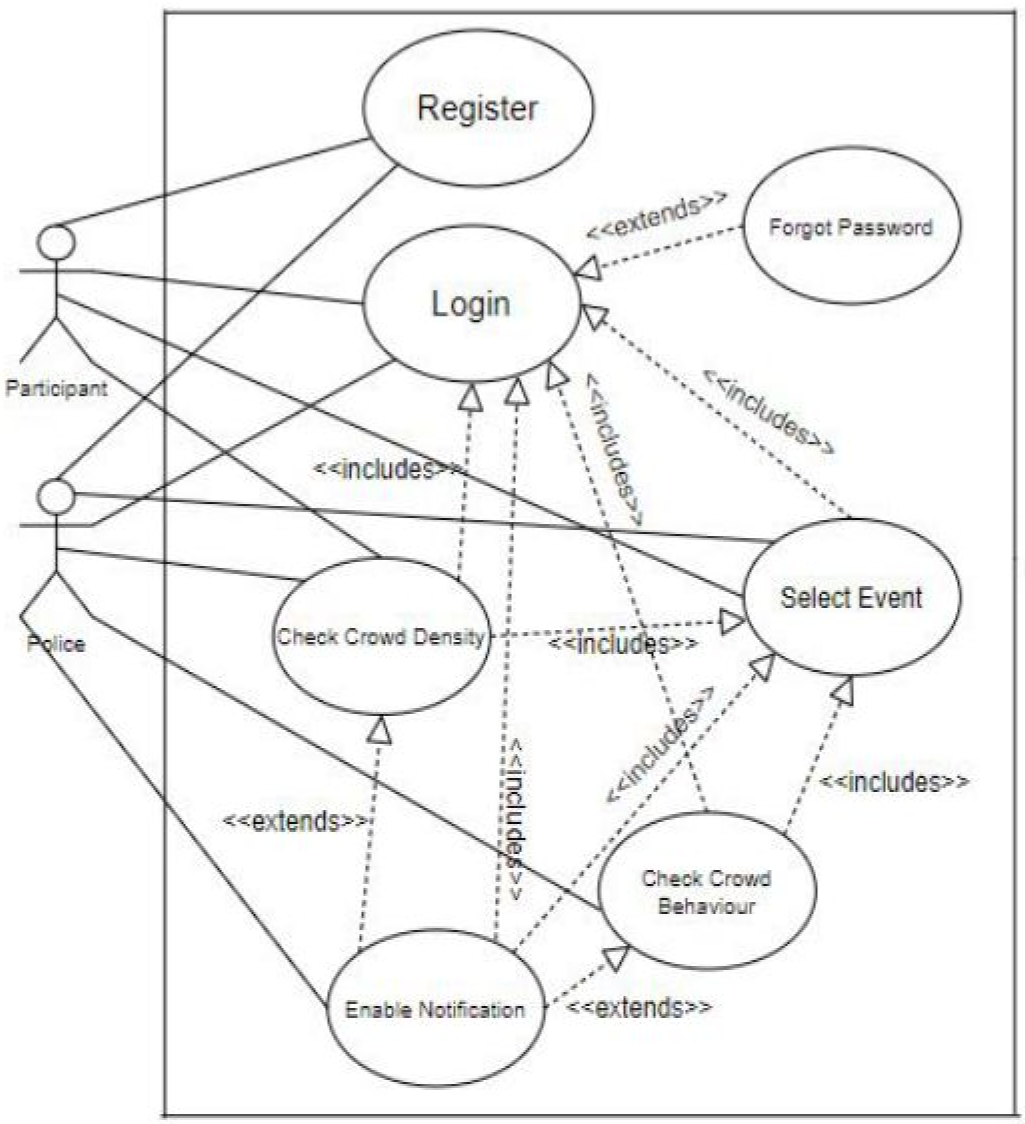


Figure 4: Use Case Diagram

4.2.3 Class Diagram

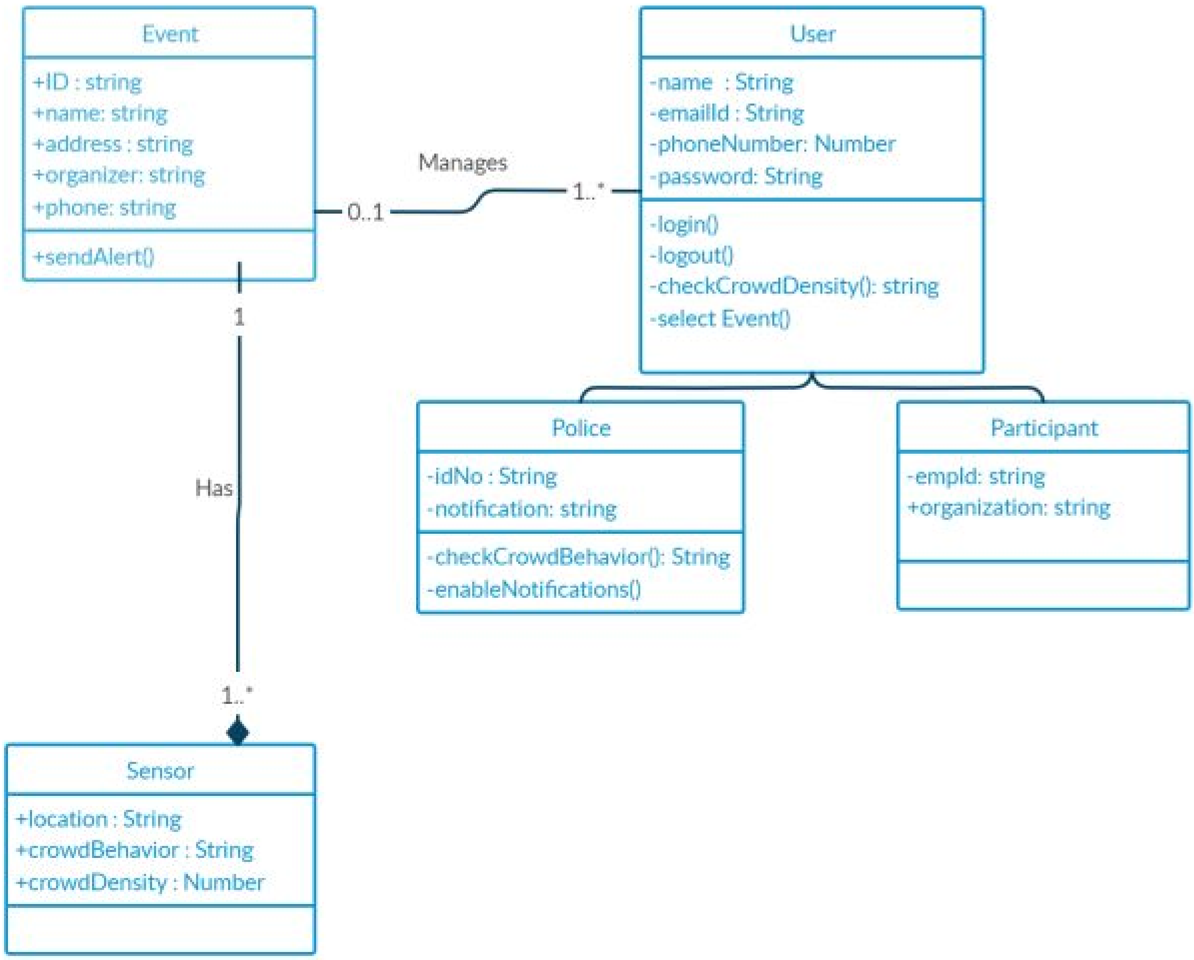


Figure 5: Class Diagram

4.2.4 Activity Diagram

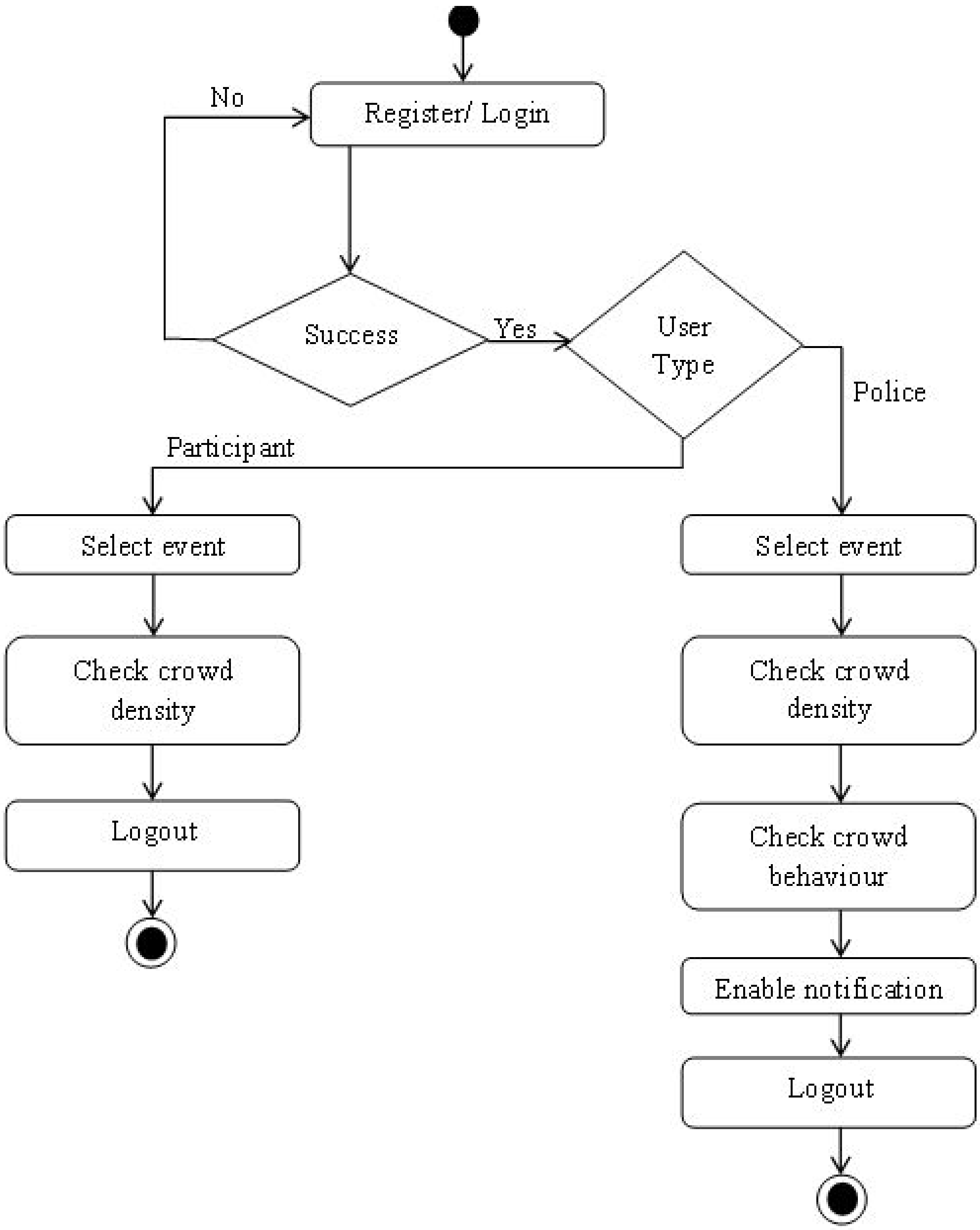


Figure 6: Activity Diagram

4.2.5 Sequence Diagram

For Police

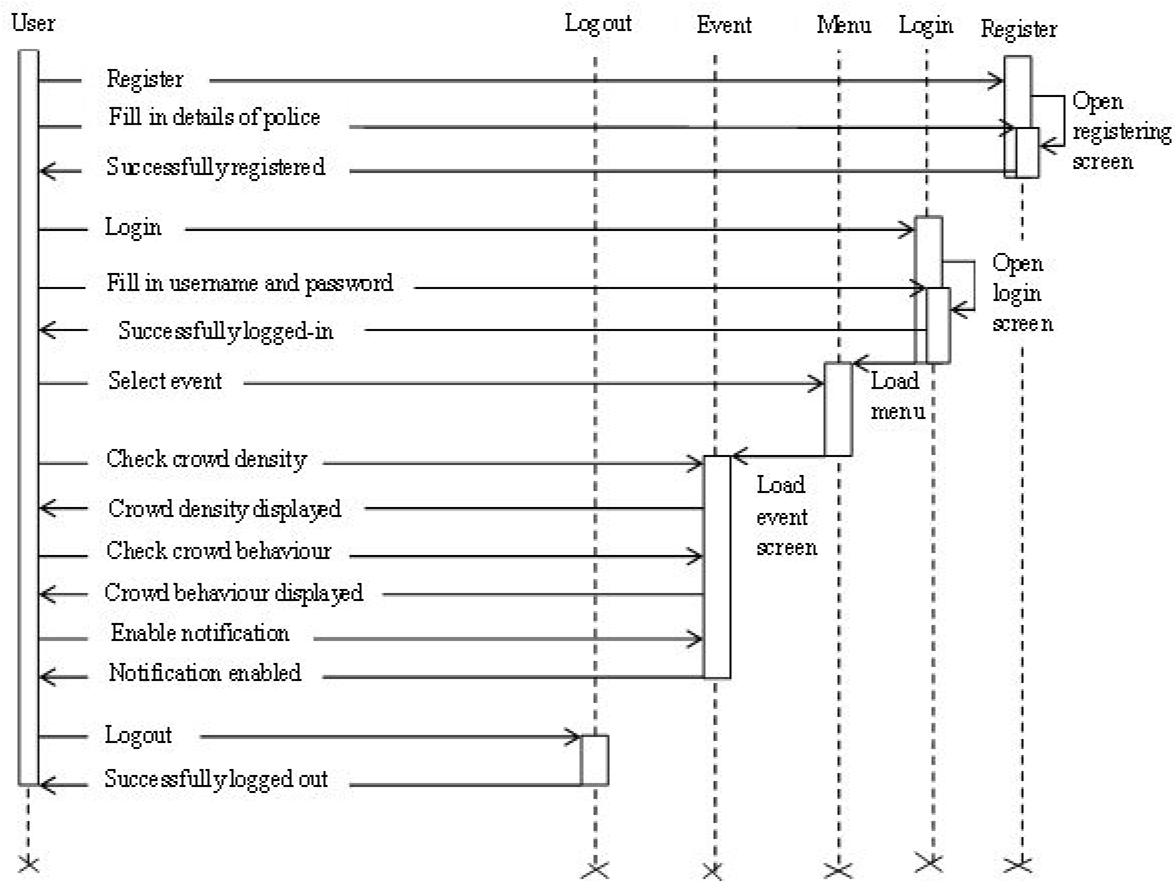


Figure 7: Sequence Diagram for Police

For Participant

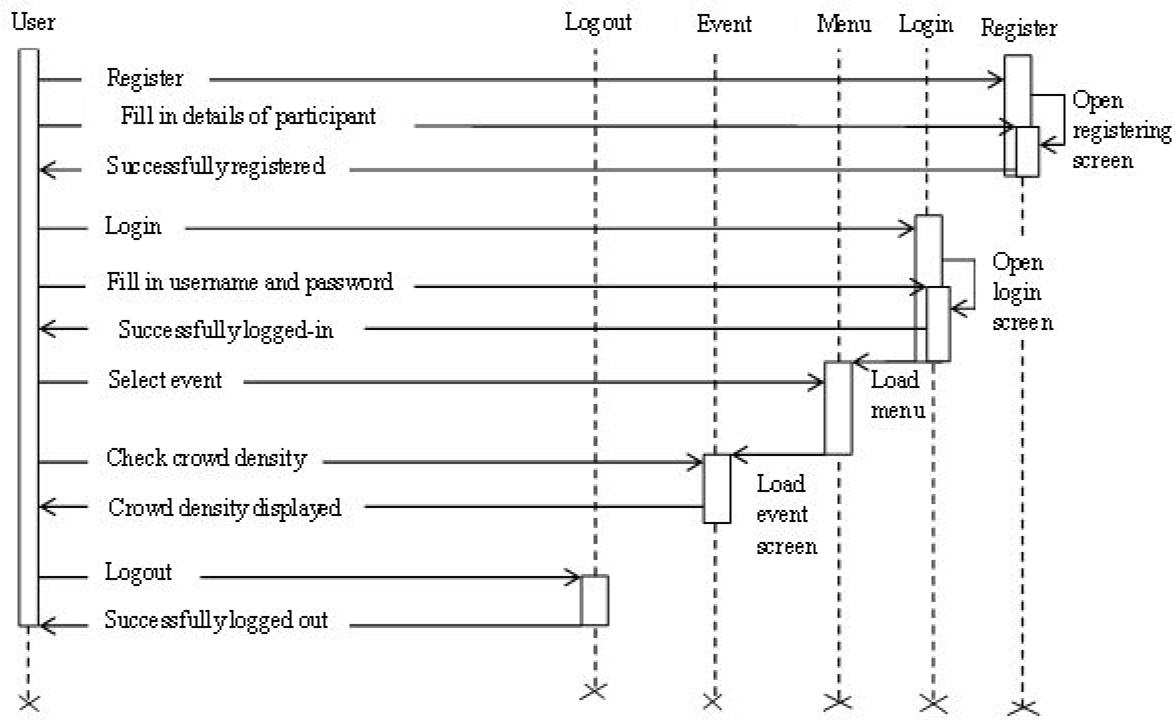


Figure 8: Sequence Diagram for Participant

4.2.6 State Chart Diagram

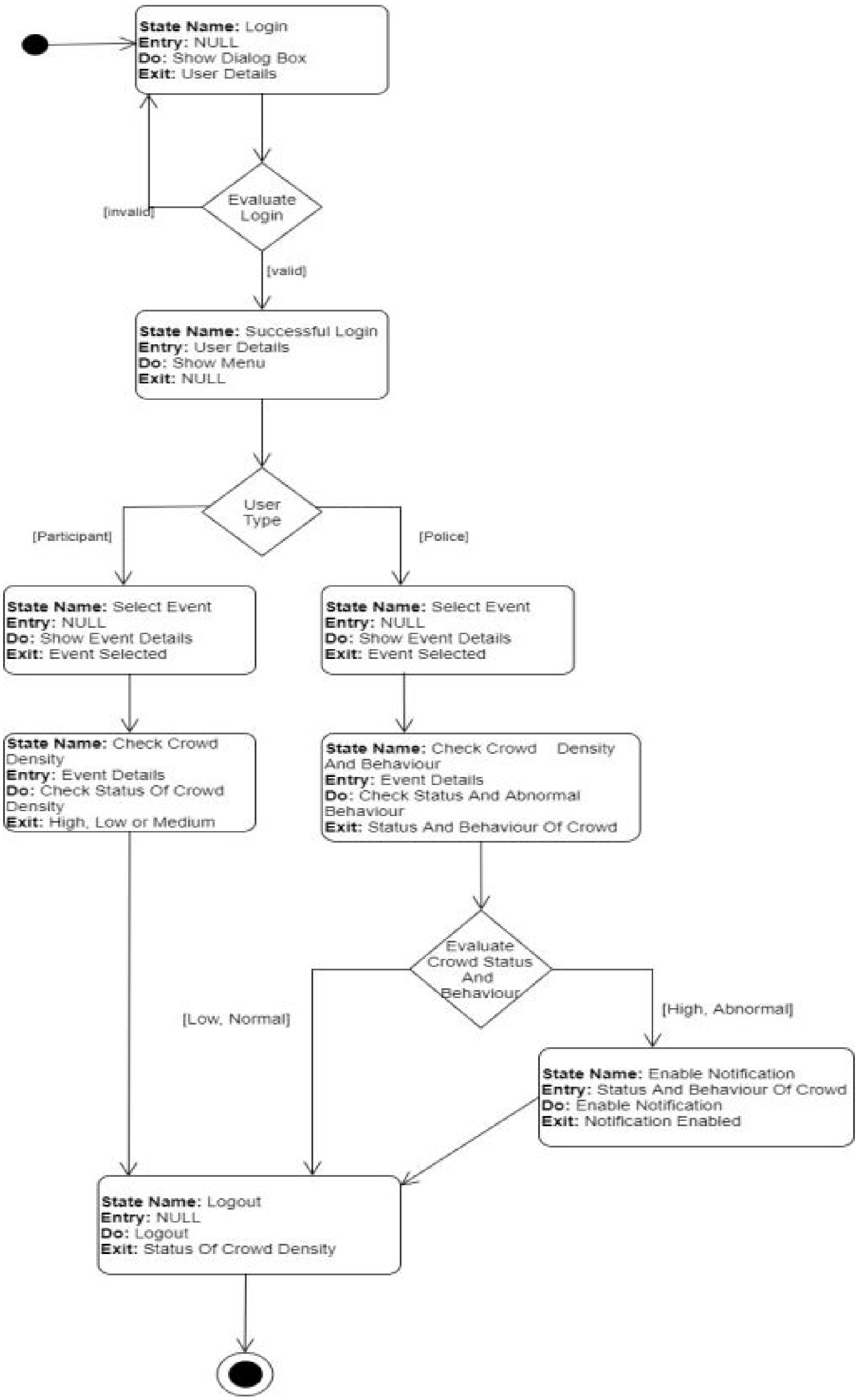


Figure 9: State Chart Diagram

4.2.7 Component Diagram

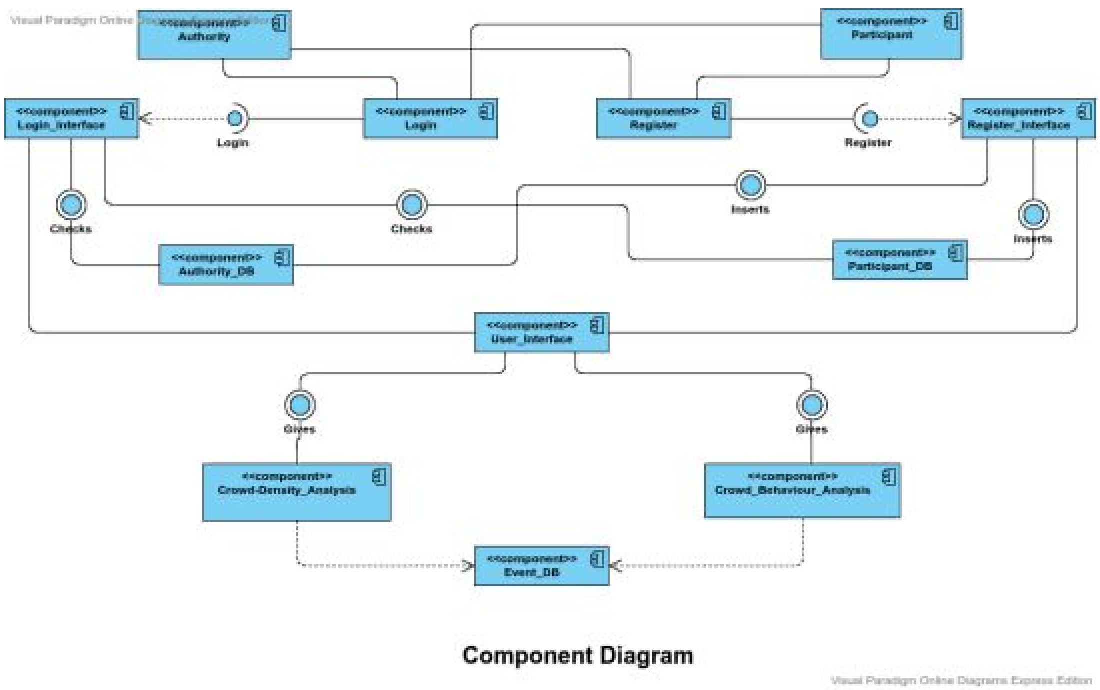


Figure 10: Component Diagram

4.3 User Interface Diagrams

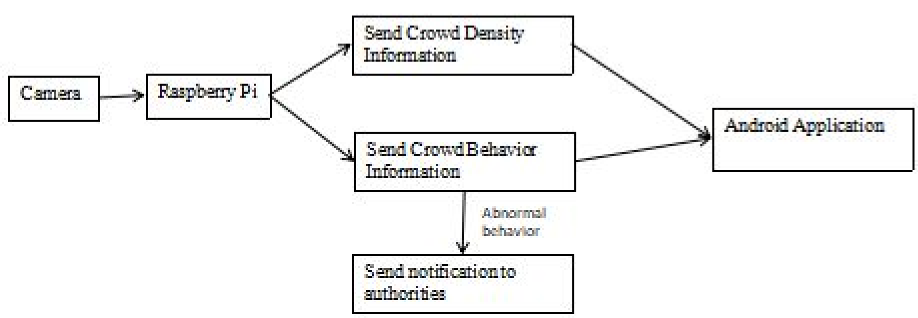


Figure 11: Interface Design Diagram

4.4 Snapshots of Working Prototype

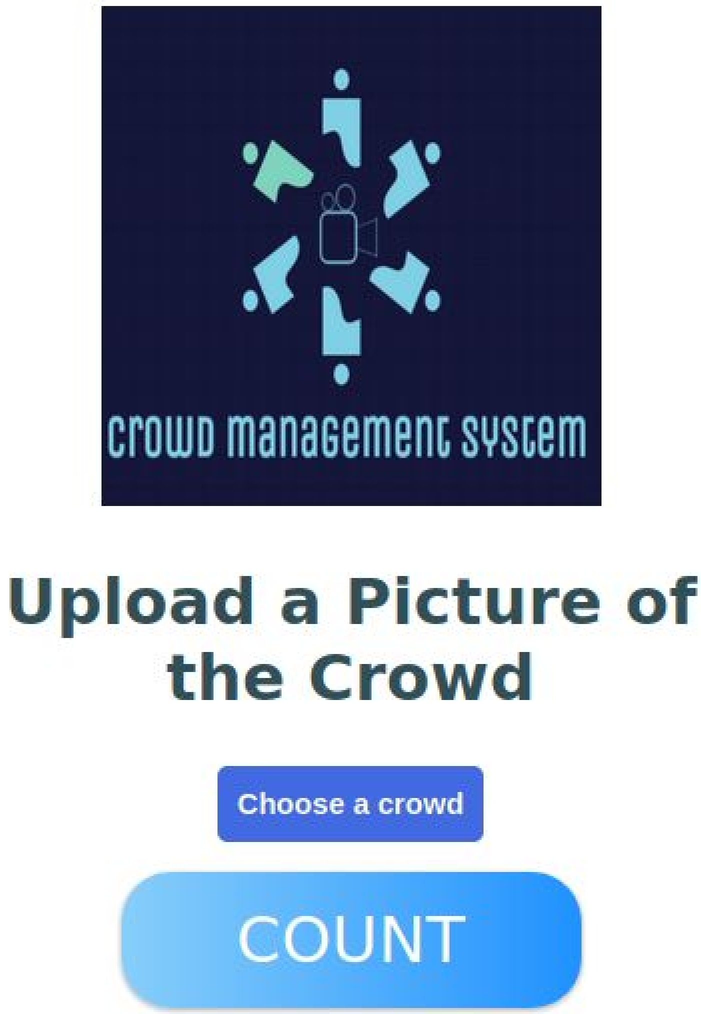


Figure 12: Snapshot 1 of Web Application

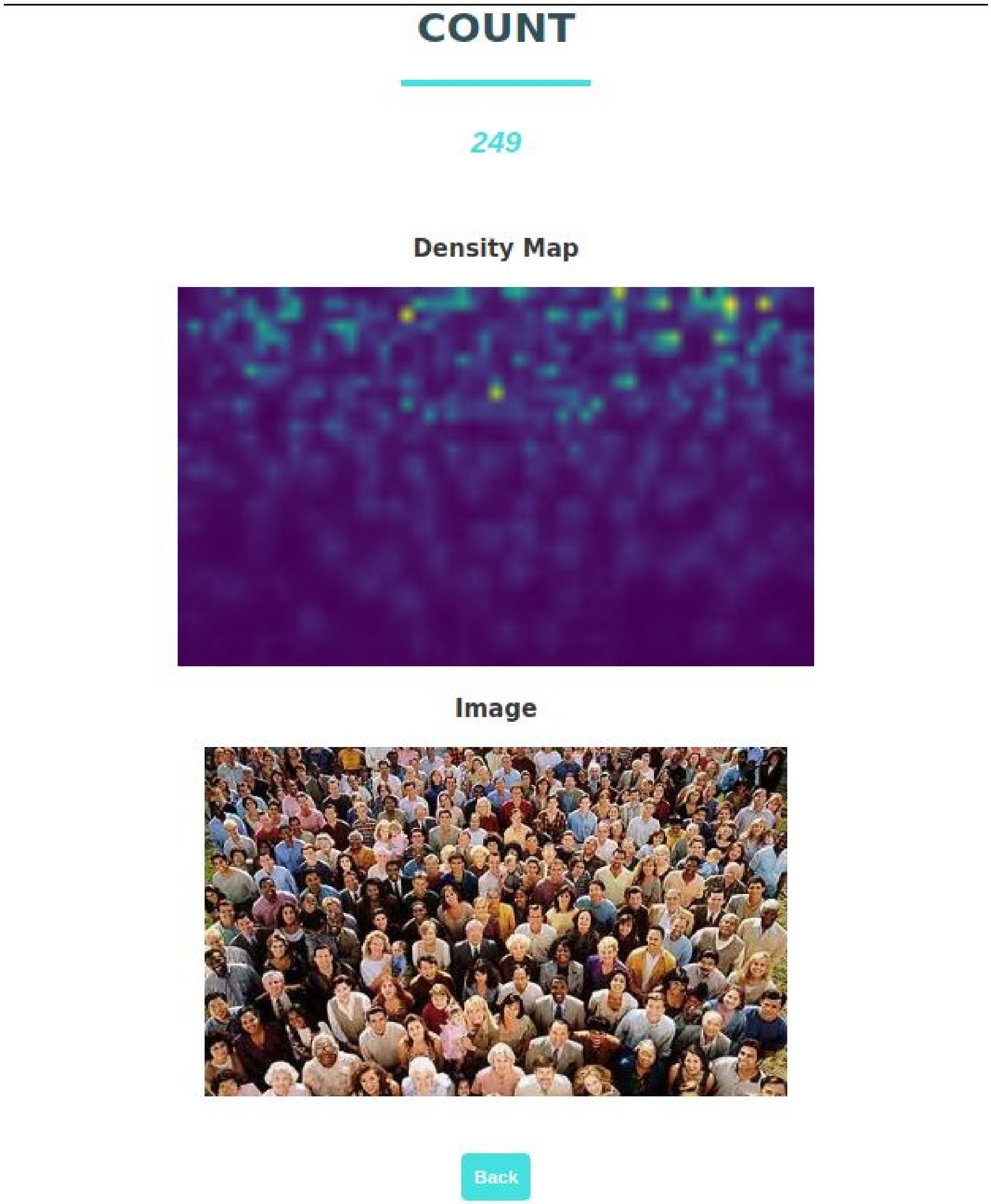


Figure 13: Snapshot 2 of Web Application

CONCLUSIONS AND FUTURE SCOPE

5.1 Work Accomplished

In accordance with the objectives, the work accomplished is as follows:

- Studied and reviewed the existing work related to crowd surveillance in public places.
- Proposed and implemented a framework for an automated image processing application for depicting the crowd count or crowd density.
- Developed a mobile application for interaction with the user.

The other two objectives are yet to be accomplished in the coming months. Those are as follows:

- To propose and implement a framework for an automated image processing application in real environment to detect crowd behaviour.
- To test and validate the system thereby develop and improve performance efficiency.

5.2 Conclusions

We have done both detecting crowd behavior and depict crowd count by using two different algorithms and then combining them. YOLO is used for crowd count and optical flow for detecting crowd behavior. We have shown how to detect both crowd count and crowd behavior by taking video as an input and developed an application in real environment to detect crowd behavior and depicting crowd count. To improve the performance efficiency we also tested and validated the system. With the help of our application any abnormal activity in the crowd can be detected and can be used for various other purposes also. We have also shown that there are many advantages of detecting crowd behavior and crowd count

5.3 Environmental Benefits

1. Economic
    - A retail store's conversion ratio can be calculated and performance can be compared across a worldwide network.
    - Events like football matches, fun fares, auditorium performances etc can be organised in a well defined and planned manner avoiding any misbehaviour or stampede.
    - Building layouts and staffing levels can be optimised.
    - Customer services in restaurants or any other businesses can be improved.

2. Social
    - Knowing the density of the crowd in a particular event, we can make healthy decisions of staying out if the crowd density is too high.
    - Violent behaviours and stampedes can be avoided.
    - Police or any other authorities can be notified well in time if any mishappening occurs during an event.
    - Safety of people attending any social gathering can be ensured.

5.4 Future Work Plan

We want to try Hyper-parameters which means number of layers, layer size, or activation functions. In our project we ran an off-the-shelf VGG-16/CSRNet structure but there are lots of possibilities for modification. Some areas that we highlighted include adjusting the learning rate and using a longer training period on an enhanced GPU. We want to test different learning rates on a logarithmic scale in the near future. Our web application now is limited because of the size of our model. One solution can be to use transfer learning to train a more efficient version of the model. Transfer Learning can be used to develop lightweight ML models. If we achieve this, we could look into publishing an Android app.

REFERENCES

[1] Lu Zhang, Miaojing Shi, Qiaobo Chen, "Crowd counting via scale-adaptive convolutional neural network".

[2] Desen Zhou, Shenghua Gao, Siqin Chen, Yi Ma, Yingying Zhang, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network"

[3] I. Laptev, J. Sivic, J.-Y. Audibert, M. Rodriguez, "Data-driven crowd analysis in videos"

[4] Sayan Burman, Agnip Mazumdar, Nilanjana Dutta Roy, "Crowd Behavior Analysis and Alert System Using Image Processing"

[5] X. Xu, S. Gong, T. M. Hospedales, "Zero-Shot Crowd Behavior Recognition" in *Group and Crowd Behavior for Computer Vision,* V. Murino, M. Cristani, S. Shah, S. Savarese, Academic Press, 2017, Pages 341-369

[6] Davies, A.C., J. Hong, Y., Velastin : "Crowd monitoring using image processing" in *Electronics & Communication Engineering Journal* S.A. 7(1), 37–47 (1995)

[7] H. Sabzevari, PAVIS, DISI, "Emotion-Based Crowd Representation For Abnormality Detection" in 2016

[8] Cem Direkoglu,Melike Sah,Noel E. O'Connor, "Abnormal Crowd Behavior Detection Using Novel Optical Flow-Based Features" in *Middle East Technical University – Northern Cyprus Campus*, Scientific Research Project (SRP) Fund (Grant no: FEN-16-YG-10).

[9] N. N. A. Sjarif, S. M. Shamsuddin, S. Z. M. Hashim, S. S. Yuhaniz, “Crowd Analysis and Its Applications” in *Communications in Computer and Information Science,* January 2011

[10] D. B. Sam, S. Surya, R. V. Babu, “Switching Convolutional Neural Network for Crowd Counting” in *Indian Institute of Science Bangalore, India*, August 2017

[11] Y. Li , X. Zhang, Deming Chen, “CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes” in *University of Illinois at Urbana-Champaign*, Beijing University of Posts and Telecommunications, April 2018

[12]. M. Gnouma, R. Ejbali, and M. Zaied, “Abnormal events’ detection in crowded scenes” in *Multimedia Tools and Applications*, 2018.

[13]. 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," at *IEEE Winter Conference on Applications of Computer Vision (WACV 2018)*, 2018.

[14]. R. V. H. M. Colque, C. A. C. Junior and W. R. Schwartz. “Histograms of optical flow orientation and magnitude to detect anomalous events in videos”in *IEEE Trans. Circuits Syst. Video Technology*.

[15] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam, “Real-Time Object Detection with Yolo”

[16]. G. Tripathi, K. Singh and D.K. Vishwakarma. “Convolutional neural networks for crowd behaviour analysis: a survey” in *Visual Computing*, 2019.

[17]. X. Wang and C.C. Loy, “Deep Learning for Scene Independent Crowd Analysis” in *Group and Crowd Behavior for Computer Vision*, 2017.

[18]. R. Mehran, A. Oyama, M. Shah. “Abnormal Crowd Behaviour Detection using Social Force Model” at *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[19]. Rodriguez, Mikel D., and Mubarak Shah. ”Detecting and segmenting humans in crowded scenes.” Proceedings of the 15th ACM international conference on Multimedia. ACM, 2007.

[20]. Zhao, Tao, and Ramakant Nevatia. ”Tracking multiple humans in complex situations.” IEEE transactions on pattern analysis and machine intelligence 26.9 (2004): 1208-1221.

[21]. M ONNET , A., MITTAL , A., P ARAGIOS , N., AND R AMESH , V.2003. Background modeling and subtraction of dynamic scenes. In IEEE International Conference on Computer Vision (ICCV). 13051312.

[22]. Orwell, James, P. M. Remagnino, and Graeme A. Jones. ”Optimal color quantization for real-time object recognition.” Real-Time Imaging 7.5(2001): 401-414