CHAPTER 1
INTRODUCTION

This chapter briefly overviews the vision-based crowd analysis (CA), a general CA framework, and discusses different CA tasks. The primary motivation behind the proposed work, an introduction to the problem statement, and the objectives of the thesis are also discussed in this chapter. Finally, the chapter concludes with a list of contributions to the thesis, followed by thesis organization.

1.1 Background

The exponential growth of the worldwide population has caused huge crowd formation in different places, for instance, rallies, stadiums, pilgrim places, public speeches, concerts, and mass transit. These places are very prone to crowd calamities. The crowd may have similar or different motion patterns. People's motion patterns and activity characterize the crowd's behavior at an abstract level. Crowd can be characterized as "Normal Crowd", "Panic Crowd", "Obstacle/Abnormal Crowd", "Congested Crowd", and "Violent/Fight Crowd". Figure 1.1 shows examples of different crowd scenes based on their behavior.

According to Rodriguez *et al.* [1], a crowd can be structured or unstructured. The structured crowd always has a common and dominant motion pattern that hardly changes over a longer period of time. On the other hand, the unstructured crowd has varying motion patterns and frequently changes over a while. Figure 1.2 shows some examples of structured and unstructured crowd scenes. The unstructured crowds are most likely prone to disasters in the presence of mobs, stampedes, violence, or any abnormal activities in the crowd. Various incidents and crowd deaths due to stampedes and violence have been recorded in past years.

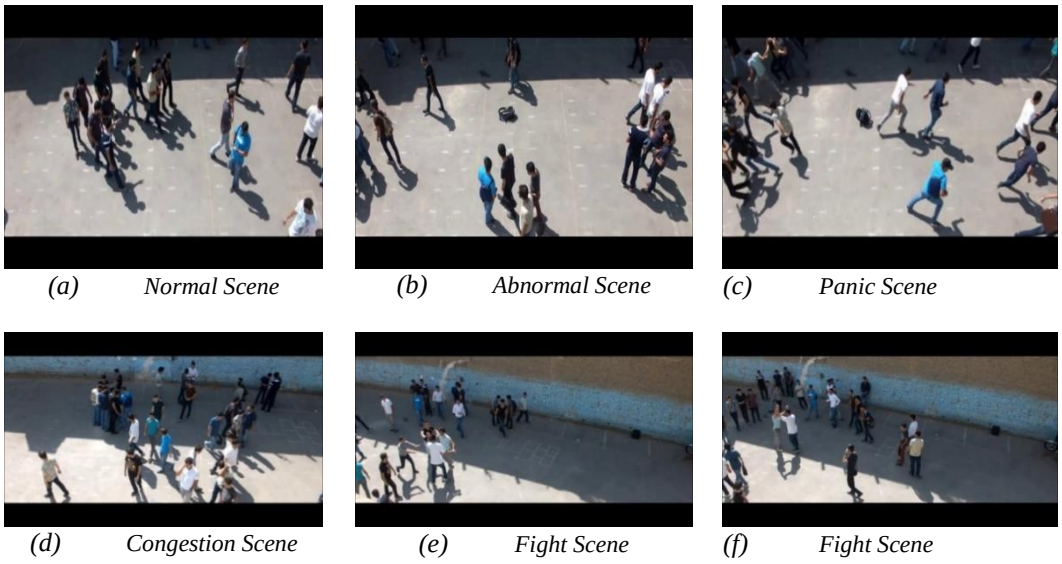

Figure 1.1: Sample of different types of crowd scenes of the MED dataset [2]

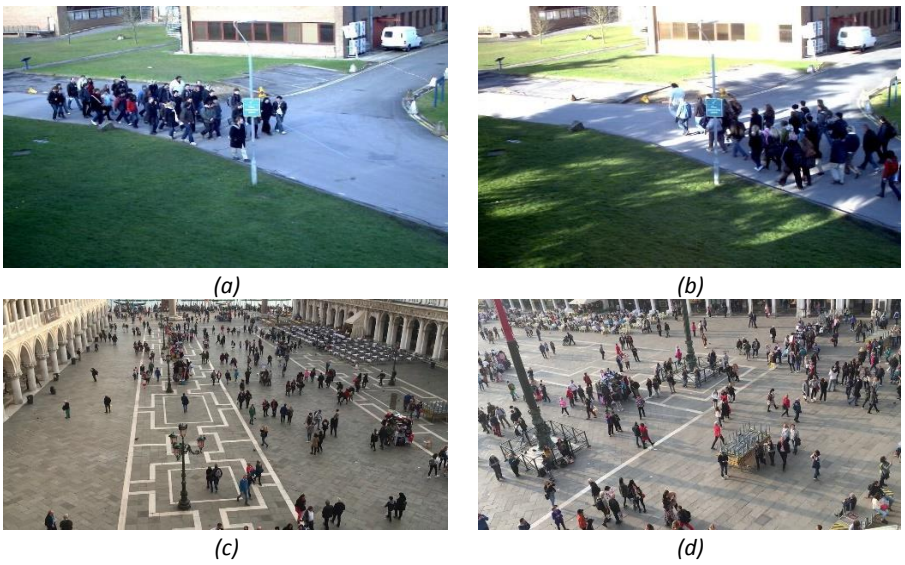

Figure 1.2: Samples of crowd scenes. (a) and (b) are the samples of the Pets 2009 datasets [3] representing structured crowd scenes. (c) and (d) are the samples of the Venice datasets [4] representing unstructured crowd scenes.

Mentioning a few, in 2008, in Jodhpur, India, a stampede in the Chamunda Devi temple killed 224 people. In a stampede in 2010, Phnom Penh killed more than 347 people. A nightclub fire in the city of Santa Maria in 2013 killed 242 people and injured 168 people. In 2015, a stampede in Mina claimed the death of more than 2400 people. A protest in Ethiopia took the life of 300 people.

A stampede in Afghanistan in 2020 caused 15 deaths. In 2021 at least 45 people were killed in a stampede in Tanzania, and many were critically wounded. Statistics of death of people due to crowd calamity from 2001 to 2022 worldwide have been made based on information available on wiki [5], shown in Figure 1.3.

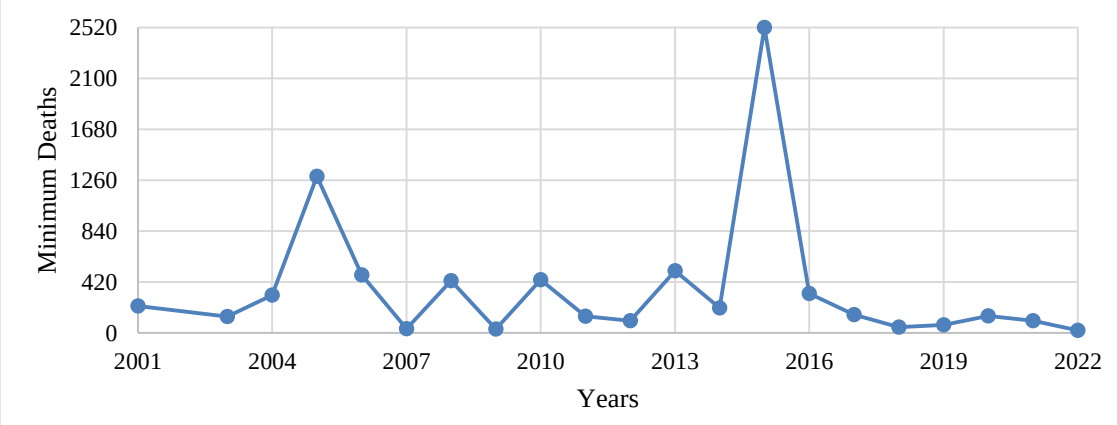

Figure 1.3: Human deaths due to stampede [5]

These calamities can be avoided by proper analysis of crowd scenes. The crowd analysis is essential to provide helpful information regarding crowd statistics and behavior, which will be very helpful in drawing effective crowd management strategies, designing public spaces, creating a virtual environment, and designing automated visual surveillance systems. The traditional approach for crowd analysis is made manually by deploying security personnel such as police, military, or authorized volunteers. The manual processes are always prone to error and are very difficult to analyze crowd behaviors, thus leading to security and safety threats. Over the last two decades, computer vision-based artificial intelligent (AI) solutions have been developed for crowd analysis which eases the process of traditional approaches. For example, a Hong Kong start-up [6] has been working on a computer vision-based AI tech project for managing crowds in universities, private businesses, and railway operators. The project's central theme is to count and track people with preserving people's privacy. Similarly, the Kumbh Mela in India, which is called to be one of the holiest and most sacred festivals, has witnessed the

first use of an AI-based visual surveillance control room to manage the crowd in 2019 [7]. Over 1000 CCTV cameras were deployed for crowd management, and it has been observed that around 250 million people participated in the Mela, and no causalities have been recorded. Thus, there is a need to facilitate computer vision-based approaches for crowd analysis

1.2 Vision-based Crowd Analysis

The computer vision solutions have a wide range of applied research areas, including human activity recognition, human pose estimation, image forgery detection, and crowd analysis. In recent years, several researchers in the computer vision community have shown a massive interest in crowd analysis. The vision-based crowd analysis is one of the essential tools for the development of smart cities and for developing efficient crowd management solutions. Figure 1.4 shows the basic structure of a vision-based crowd analysis system and its possible applications.

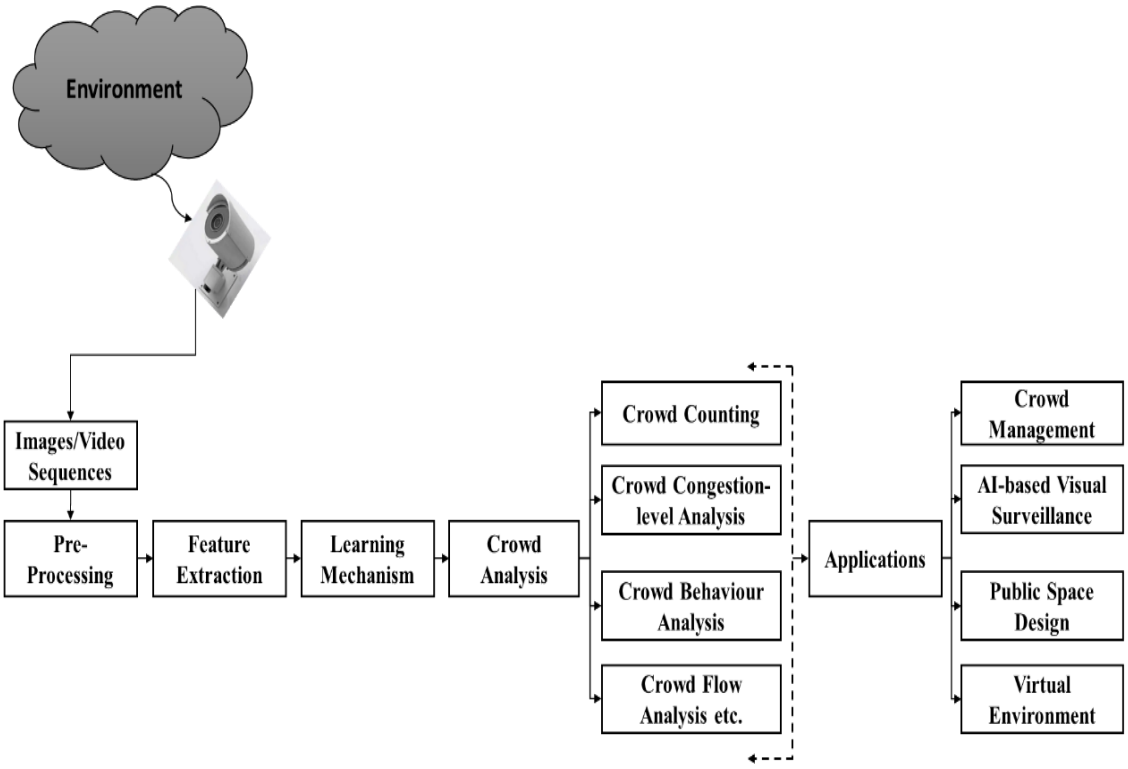

Figure 1.4: Overall structure of vision-based crowd analysis system and its applications

The vision-based crowd analysis system acquires video sequences or still images from the camera. The acquired data are often pre-processed. The pre-processing may include image/frame resize, color conversion, brightness or contrast enhancement, and noise removal. The pre-processed data is then used to extract meaningful features. These features are used by some learning mechanisms to perform different crowd analysis tasks like crowd counting, crowd congestion-level analysis, crowd behavior analysis, and crowd tracking. The valuable information drawn from crowd analysis can be applied to crowd management, public space design, virtual environment creation, and AI-based visual surveillance.

Most of the existing solutions for crowd analysis are based on a single or auxiliary task. The performance of any crowd analysis task mainly depends on two critical steps, i.e., feature representation and learning mechanism. Many techniques using handcrafted and deep learning approaches have been developed for feature representation. Handcrafted features like local binary patterns [8], Histogram of Oriented Gradients (HOG)[8], Histogram of Optical Flow (HOF) [9], trajectories [10], spatial-temporal interest points [11] have been vastly explored to represent crowd scenes. However, such feature representation requires high domain knowledge. Such limitation is overcome by automatic feature extraction using deep learning techniques. State-of-the-art deep models like CNN [12], encoder-decoders [10], LSTM [11], Conv-LSTM [12] and GANs[13] have been used for automatic feature modelling for crowd scenes.

Different learning mechanisms like supervised [13], unsupervised [14], weakly supervised [15], and reinforcement methodologies [16] have been used for different crowd analysis tasks. Most of the crowd counting approaches follow supervised and weakly supervised-based regression strategies. Whereas supervised and unsupervised solutions can be found for crowd behavior analysis and crowd tracking or trajectory

predictions. Recently few re-enforcement learning approaches have been developed for crowd counting. However, deep learning mechanisms are more effective than conventional machine learning approaches.

Out of different tasks of the CA, the minimum information required to control crowd disasters is the crowd density, crowd congestion, and crowd behavior. Also, utilizing several models for the CA will incur computational and synchronization issues, which can be overcome by developing a multitasking model. Hence, this work focuses on the significant tasks of crowd analysis: (i) crowd counting and density estimation, (ii) crowd congestion-level analysis, (iii) crowd behavior analysis, and (iv) multitasking CA, which provides the essential information required for crowd management, public space design, and developing a visual surveillance system thereby minimize crowd disasters.

1.2.1 Crowd Analysis Tasks

1.2.1.1 Crowd Counting and Density Estimation

Crowd counting [17] means the number of people present in a scenario, whereas density estimation, formally crowd density estimation [18], refers to a technique used for crowd counting. Here, the crowd counting is done by performing regression on the ground-truth crowd density maps. The crowd count becomes helpful information for evacuation, monitoring, and designing public spaces.

1.2.1.2 Crowd Congestion-Level Analysis

The crowd congestion-level analysis [19] provides valuable information regarding different congestion levels of the crowd. For example, very-low, low, medium, high and very-high are the terminologies used for crowd congestion labeling. Such analysis is handy in crowd monitoring and avoiding any panic-like situations.

1.2.1.3 Crowd Behaviour Analysis

Crowd behavior analysis (CBA) [11] analyzes the appearance or motion patterns of different crowd sequences for predicting crowd behaviors such as normal, fight, panic, or anomalous (abnormal) crowd activities. A compelling visual surveillance system requires efficient crowd behavior analysis to provide security and safety.

1.2.1.4 Multitasking Crowd Analysis

The multitasking CA [14], [15] deals with performing several crowd analysis tasks using an underlying machine learning or deep learning model. The main objective behind the design of such type of model is to minimize the complexity of the use of multiple single task-based CA models and also to avoid synchronization issues during decision making.

1.2.2 Need for Vision-based Crowd Analysis

The vision-based crowd analysis provides valuable information, which helps to avoid crowd calamities. There are some prominent reasons which demand the need for vision-based crowd analysis.

- Minimizing the human errors caused during crowd analysis.
- Improved visual surveillance for crowd monitoring and security surveillance.
- Efficient crowd management strategies.
- Minimizing the crowd disaster.

1.3 Motivation

Public domains such as stadiums, pilgrim places, religious places, protests, marathons, rallies, shopping malls, and public transport are open to forming large crowd gatherings. These places have potential risks of crowd hazards like stampedes and

violence. Such disasters result in massive loss of lives and public or private properties. Therefore, proper crowd analysis is required to acquire helpful information to avoid crowd disasters, providing a secure and safe environment for society. The major crowd analysis tasks can provide essential information that helps develop better crowd management strategies, better design public spaces, and develop an intelligent surveillance system, thus reducing crowd casualties.

A crowd has both psychological characteristics and dynamic behavior, making the crowd analysis a tedious task. Various traditional and computer vision approaches have been developed for crowd analysis. The traditional approaches are manual processes and perform well in low or moderate crowd densities but not dense crowd scenes. With the advancement of computer vision and machine learning, such drawbacks can be overcome. As shown in Figure 1.4, the vision-based crowd analysis has three crucial stages: pre-processing, feature representation, and learning mechanism. Different solutions have been proposed for each stage, but the performance of the crowd analysis model is always affected by occlusion, cluttered background, scale variation, and illumination changes in the crowd scene.

Hence, such vital issues need to be resolved by proposing new computer vision and machine learning techniques to find better pre-processing operations, feature descriptors, and learning mechanisms, thereby improving the performance of main tasks of crowd analysis: crowd counting and density estimation, crowd congestion-level analysis, crowd behavior analysis, and multitasking CA.

1.4 Problem Statement

This thesis addresses some of the critical challenges and issues of different crowd analysis tasks using computer and machine learning approaches. The problem statement of the thesis is:

"Design and development of some methods and models for crowd analysis using computer vision and deep learning approaches.

1.5 Thesis Objectives

The objectives of this thesis are,

I. Conducting a comprehensive literature review of state-of-the-art methods, their comparative analysis (CA), and identifying research gaps in different tasks of CA such as crowd counting and density estimation (CCDE), crowd congestion-level analysis (CCA), crowd behavior analysis (CBA), and multitasking CA. Further, the objective of the literature review is to briefly discuss the datasets and performance metrics used for experiment and results analysis.

II. Design and development of two novel models using deep learning techniques for the task of video-based crowd counting by addressing the research gaps identified in the literature review of crowd counting and density estimation.

III. Design and development of a novel crowd congestion-level analysis model using deep learning approaches by fulfilling the research gaps identified in the literature review of crowd congestion-level analysis.

IV. Design and development of two novel architectures for crowd behavior analysis (CBA) using conventional machine learning and deep learning techniques. The first architecture under the contributions of CBA is the design of a one-class classification (OCC)-based approach using conventional machine learning and deep learning techniques for crowd panic detection. On the contrary, the second deep architecture is based on the design of a multi-class classification (MCC)-based approach using deep learning techniques. Both the models are designed to fulfill the research gaps identified for the task of CBA.

V. Design and development of a novel multitasking crowd analysis (CA) model using deep learning concepts to address the identified research gaps in the task of multitasking CA. Multitasking concentrates on crowd counting and crowd behavior prediction. Another objective under this contribution is to develop a largescale multitasking CA dataset using publicly available crowd behavior datasets, i.e., MED and GTA, to fulfill the multitasking CA dataset requirement.

1.6 Contributions to the Thesis

The contributions to the thesis are summarized as follows,

I. Performed a comprehensive literature study of state-of-the-art methods for significant tasks of crowd analysis: crowd counting and density estimation (CCDE), crowd congestion-level analysis (CCA), crowd behavior analysis (CBA), and multitasking CA to identify the research gaps.

II. Proposed two novel deep architectures for video-based CCDE. The first model is an attentive multi-stream CNN architecture, whereas the second is a cascaded deep architecture with weak supervision designed for video-based CCDE to fulfill the identified research gaps.

III. Proposed a real-time crowd congestion-level analysis model using two input stream multi-column multi-stage CNN architecture.

IV. Proposed two novel architectures for crowd behavior analysis using conventional machine learning and deep learning approaches. The first model utilized conventional machine learning and deep learning concepts and designed a multiscale spatial-temporal 3D atrous-net and PCA-guided OC-SVM designed to fulfil the identified research gaps for crowd panic detection. On the other hand, the proposed second model is for crowd behavior classification designed using a two-stream multiscale deep architecture.

V. Proposed a multitasking crowd analysis model using multiscale flow attentive depth separable CNN. Also, a largescale multitasking CA dataset is prepared using the standard benchmark crowd behavior datasets: the MED, the GTA.

1.7 Thesis Organization

The thesis is organized into seven chapters as follows,

Chapter 1 briefly discusses the background of crowd analysis and the need for vision-based crowd analysis. Further, motivation for the research, problem statement, and thesis objectives are also explained. This chapter also highlights main contributions to the thesis and presents thesis organisation.

Chapter 2 presents a brief literature review of state-of-the-art conventional and deep learning-based approaches for crowd analysis tasks such as Crowd Counting and Density Estimation (CCDE), Crowd Congestion-level Analysis (CCA), Crowd Behavior Analysis (CBA), and multitasking CA. This chapter also discusses the details of the datasets and performance metrics used in the thesis.

Chapter 3 presents two novel deep architectures for video-based CCDE by addressing the identified research gaps. The two proposed models are: "AMS-CNN: Attentive Multi-Stream CNN for Video-based Crowd Counting" and "A Novel Cascaded Deep Architecture for Video Crowd Counting with Weakly-Supervised Learning." This chapter also demonstrates experiments on the publicly available datasets and comparative results analysis with state-of-the-art methods available in the literature. The proposed models have been observed to perform better than the state-of-the-art approaches mentioned in the literature.

Chapter 4 discusses a proposed deep architecture for crowd congestion-level analysis. This chapter briefly discusses the background of this work and also presents a detailed working of the proposed model entitled "A Real-time Two Input Stream Multi-

Column Multi-Scale CNN (TIS-MCMS-CNN) for Efficient Crowd Congestion-level Analysis." This chapter also discusses the detail of the creation of a crowd congestion-level dataset. The experiment details and comparative results analysis with state-of-the-art methods have been discussed. The proposed models have been observed to perform better than the state-of-the-art approaches mentioned in the literature.

Chapter 5 presents and discusses two proposed models for crowd behavior analysis which are "MuST-POS: Multiscale Spatial-Temporal 3D Atrous-net and PCA guided OC-SVM for crowd panic detection" and "TS-MDA: Two-Stream Multiscale Deep Architecture for Crowd Behavior Prediction." This chapter briefly discusses the background details and research gaps identified for the crowd behavior analysis. The chapter explains experiment details and comparative results analysis with state-of-the-art methods available in the literature. It has been observed that the proposed models perform better than the state-of-the-art approaches mentioned in the literature.

Chapter 6 presents and discusses a proposed multitasking CA model i.e., "Multiscale Flow Attentive Depth Separable CNN for Multitasking Crowd Analysis." This chapter briefly explain the background details and research gaps behind multitasking CA. This chapter also provides a detailed explanation behind creating the multitasking crowd analysis dataset. The chapter explains experiment details and comparative results analysis with state-of-the-art methods available in the literature. The proposed models have been observed to perform better than the state-of-the-art approaches mentioned in the literature.

Chapter 7 summarizes the research findings and concludes the thesis. This chapter also discusses some of the future research scopes in crowd analysis.