**Scalable Multi-Model Deep Ensemble Learning Framework for Mob Behavior Analysis**

A Research Proposal

Presented

by

Contents

[ABSTRACT 3](#_gjdgxs)

[1. Rationale and Significance of the Study 4](#_1fob9te)

[1.1 Background 4](#_3znysh7)

[1.2 Introduction 5](#_2et92p0)

[1.3 Research Objectives 6](#_tyjcwt)

[1.4 Scope 6](#_3dy6vkm)

[1.5 Problem Statement 7](#_1t3h5sf)

[2. Related Research Work 8](#_4d34og8)

[2.1 Literature Survey 8](#_2s8eyo1)

[2.1.1 Crowd Counting 12](#_17dp8vu)

[2.1.2 Crowd tracking 14](#_3rdcrjn)

[2.1.3 Crowd Abnormality Detection 16](#_26in1rg)

[2.2 Research Gap 17](#_lnxbz9)

[3. Proposed Research Methodology 18](#_35nkun2)

[3.1 Proposed Framework 18](#_1ksv4uv)

[3.2 Dataset 19](#_44sinio)

[4. Evaluation matrix 21](#_2jxsxqh)

[4.1 Precision 21](#_z337ya)

[4.2 Recall 21](#_3j2qqm3)

[References 22](#_1y810tw)

# ABSTRACT

Manual monitoring of mob analysis from the surveillance cameras requires continuous presence of the human supervisor. This manual supervision is very challenging because there will be multiple cameras and the human can miss some particular happenings. Due to lack of timely information regarding mob build up, law enforcement agencies become active very late, meanwhile, the mob blocks the roads, spreads widely and becomes violent. Moreover, Early and real-time information regarding mob buildup will help to manage it in good manner, and with real-time road blockage, traffic police can make traffic diversion plans. Although efforts had been made to automate the mob analysis process, each work had touched on a single aspect of the mob-like density, counting but lacking in determining the mob intent with prohibited objects. This work has devised an ensemble that assesses the mob buildup, nature of the mob and behavior analysis of the participants using deep ensemble features.

# **1. RATIONALE AND SIGNIFICANCE OF THE STUDY**

## 1.1 Background

In many cities today, the world's overpopulation causes crowded conditions. Parades, station exits and entries, political protests, and strikes all contribute to these packed conditions. These circumstances imply a proliferation of security concerns. The installation of surveillance systems based on video-protection cameras is happening concurrently in more and more cities. These surveillance systems were initially watched over by human operatives. However, this approach quickly proved to be cumbersome, ineffective, and prone to mistakes [1].

Smart cities have begun to appear during the past few decades. A smart city involves the application of technology to improve the quality of life for urban residents. The popularity of this idea is on par with the adoption of intelligent surveillance systems, which use autonomous systems instead of large-scale human intervention. This group of systems is used in crowd analysis. Crowd analysis is becoming more and more popular within the field of computer vision. Security forces are extremely concerned with comprehending the crowd mechanics that explain what could be dangerous in large gatherings. To comprehend how people and crowds behave in crowded settings, numerous research has been done. Crowd analysis is divided into two research axes: crowd statistics and crowd behavior analysis, according to numerous surveys. The goal of crowd statistics is to calculate crowd density using techniques for counting people in a crowd. The Level of Service (LOS) of a crowd is the most appropriate metric for measuring crowd density because it is drawn from the realm of motor traffic flow. Crowd behavior analysis is used to examine a crowd's behavior. Crowd tracking and activity analysis are the two main subfields that make up this field.

## 1.2 Introduction

Video surveillance has become an increasingly important topic as advanced technology and the demand for security. The automatic detection and analysis of abnormal events in public locations or open events are at the center of the discussion. Public assemblies, sporting events, demonstrations (e.g., strikes, protests), and other events with large numbers of people (mobs) are the main public security challenges. The safety of public events involving huge crowds has long been a top priority for relevant authorities due to the significant risk of degeneration. To meet the demand, several security firms specializing in mob control have sprung up in the past. This issue has drawn the scientific community’s attention, particularly for the automatic detection of abnormal mob behaviors on national occasions [5]. Everyone is becoming more familiar with technological platforms to manage various security responsibilities at this moment. With the affordability of CCTV and the ease with which it can be set up, it is becoming increasingly popular. These surveillance systems will not provide security unless many trained people with strong attention skills are available to watch flicks. [6],[7].

The monitoring system comprises a huge number of cameras that monitor various locations. It detects any observable, undesirable, or unlawful conduct to respond quickly and effectively to the situation. Searching for a specific activity from many recorded video files can be challenging and time-consuming when investigating an incident [8] [2]. To overcome these challenges, we need an advanced and intelligent computer vision-based system. This research work is proposed here to overcome the lack of traditional surveillance approaches. This work will help detect aberrant hu- man behavior in public, such as violent mob behavior, peaceful mob behavior, etc. In this research work, we will utilize object detection approaches to detect mobs on-road and objects like guns and sticks. In the second module, the traffic blockage will be determined based on the mob presence on the road. will give real-time information to authorities at various stages, including native police stations and police headquarters. This component aids in the early resolution of any severe effects of public violence. The entire framework contributes to successful law enforcement through smart policing

## 1.3 Research Objectives

* To classify scenes in live streaming for mob buildup (e.g., small, middle, and large).
* Mob analysis makes predictions for potential behaviors like growth, dispersion, peacefulness, and restlessness.
* Violation assessment based on prohibited objects (e.g., weapons, sticks).
* Identification of road blockage by the mob.

## 1.4 Scope

1. Following objects will be detected for behavior analysis:

* Person
* Weapon
* Sticks
* Placards
* Fire

1. Road segmentation for road blockage
2. Behavior prediction system

There is a growing trend of smart cities worldwide, where the security situation is monitored through surveillance cameras. These cameras are smart and intelligent enough nowadays. Proposed automation work lacks a unified framework for the mob analysis that can jointly analyze the statistical side and behavior analysis. This work is to devise a framework that assesses the mob buildup, nature of the mob, and behavior analysis of the participants using deep ensemble features.

## 1.5 Problem Statement

There is a growing trend of smart cities worldwide, where the security situation is monitored through surveillance cameras. These cameras are smart and intelligent enough nowadays. Proposed automation work lacks a unified framework for the mob analysis that can jointly analyze the statistical side and behavior analysis. This work is to devise a framework that assesses the mob buildup, nature of the mob, and behavior analysis of the participants using deep ensemble features.

# 2. Related Research Work

## 2.1 Literature Survey

The most recent research concentrates on mob segmentation, mob counting, mob tracking, and mob behavior analysis. Nonetheless, there is a gap in their coverage of detecting mob nature with restricted objects in their hands. However, the improvement is strongly reliant on the providing of large-scale public mob datasets to continue. The majority of the approaches that have been proposed up until this point for analyzing crowds are dependent on the description of the scene, and as a result, they can only be tested on settings that are very similar to one another.

Nevertheless, the crowd’s notions, structure, and creation are completely different from one another. Analyzing the behavior of crowds leads to similar findings, with the main difference being that the learned models are totally reliant on the training data given to them. The various crowd structures all adhere to the same underlying principles and can be categorized according to general characteristics. The ability to automatically understand such characteristics across multiple views in clips has significant apps, such as video retrieval and crowd event detection, as well as benefits in the development of crowd models that are expected to generalize to new scenes that are not in the training set. The availability of large-scale crowd datasets that cover a wide variety of scenes and video sequences is extremely important to the progress made. Let’s have a look at the research regarding mobs. This comprehensive research focuses on analyzing human motion based on techniques used in computer vision. It has reviewed three major issues in human motion analysis systems, including detection, tracking, and activity understanding of humans. These are all important aspects of human motion analysis. Although a significant amount of work in crowd analysis is described, several challenges, including segmentation, modeling, and occlusion management, are not treated in detail concerning many of the critical key characteristics stated above. This is a limitation and issue of this current work [9]. Let’s see some more research work about major issues of mass detection.

This study focuses on the most up-to-date approaches available for addressing major issues such as detection, tracking, interpretation, description of activities, and human identification for visual and interactive surveillance employing numerous cameras. Occlusion management, the fusion of two-dimensional and three- dimensional monitoring, visual surveillance, and human recognition are also included. In addition, there is no consideration given to the detection of anomalies, the prediction of behavior, and the fusion of data from many cameras and remote monitoring [10]. We have explained some other research work about computer vision-based detection of crowds and then analysis of them. This work aims to survey the various approaches to computer vision-based crowd analysis. In addition to this, specifics of the in-depth research on crowd modeling and analysis from the perspectives of sociology, psychology, and computer graphics are pro- vided. There is currently no plan to develop intelligent systems that integrate the results of non-vision-based methods with those of computer vision-based methods [11]. This article presents a survey of crowd analysis with computer vision techniques, including topics such as people monitoring, crowd density estimates, validation, simulation, and event detection. This article does not have enough time to explore all of the many online datasets and benchmarks for crowd analysis [12]. Crowd monitoring and analysis need a recognition phase as well in this process of mob analysis.

In the context of transit surveillance, this assessment focuses on human behavior recognition approaches, including single person, multiple person, person–facility, and person–vehicle interactions. There has been no discussion of the standard assessment techniques and algorithms for human behavior recognition [13]. This review of the state-of-the-art in crowd analysis from the years 2000 to 2011 covers the procedures of preprocessing, object tracking, and event/behavior detection. The problems with crowd modeling and crowd scene analysis have not been solved [14]. An overview of contemporary techniques for anomaly detection in automated surveillance, focusing on five factors including target, kinds of sensors utilized and the feature extraction procedures, anomaly definitions, learning, and modeling algorithms, is presented. The use of a variety of sensors, fields of view, and resolutions at the outside ranges of the scale and algorithms applicable to a large variety of different targets is not addressed [15]. In crowd analysis, the most important things are human behaviors. With the help of crowd behavior, we can draw useful conclusions and information.

This study has provided an overview of the many stages of human behavior analysis, such as an abstraction, a degree of semantics, and a time-oriented categorization. In addition, the feature categories, recognition strategies, and system design procedures that are most commonly employed and show the most promise have been broken down. In addition, analyses of vision and multimodal-based method- ologies were conducted for this survey. A debate on the design and development of stable systems for behavioral analysis is lacking because most systems can only tackle certain issues in constrained contexts [16]. This study focused on the se- mantically enhanced analysis of human behavior. The techniques are primarily classified by the semantics involved, thereby dividing the entire process into high- level and low-level components to better understand the techniques considered to be state-of-the-art. When it is determined that the backdrop scene is inside rather than outside, the methods used to manage multiple item interactions, both individual and group, are not considered [17]. An analysis of human behavior from the years 2000 to 2014, broken down into three primary categories: detection methods, datasets, and implementations. In addition to this, it specifies the many application areas, such as the detection of humans and pedestrians, anomalous activity, and action identification. This review does not cover the following topics: developing a framework that is less dependent on non-realistic assumptions; the fusion of various image descriptors; automatic tracking and recognition for various scenarios including clothing, clutter, and varying body shapes; and behavior representation in dynamic scenes [18]. Human behavior is important as well as human actions, which are important too.

This study examines, classifies, and summarizes current published and recognized research developments on human action recognition and provides an overview of the main processing framework of human activity recognition systems. Methods that are now considered state-of-the-art in each essential issue are explained, with a primary emphasis on three major tasks: human detection, tracking, and activity identification or behavior comprehension. In addition to this, it gives a brief description of the datasets themselves and a comparison of the characteristics shared by each of the datasets, and it discusses the publicly available datasets to conduct uniform testing of the technique proposed by the authors. The recognition of human motions from various viewing angles has not been revealed, which is an extremely crucial and significant subject that has to be considered [19].

### 2.1.1 Crowd Counting

It offers an overview of the most recent developments in approaches for the study of crowded scenes, including topics such as the learning of crowd motion patterns, the examination of crowd behavior, and the detection of crowds. Excluded from consideration for crowd analysis are real-time processing and generalization over a variety of datasets and multi-sensor information. Recent developments and emerging trends in crowd analysis are discussed within the Tracking-Learning-detection framework, Deep Learning for crowd scene analysis, and the framework of crowd modeling studies. It has been established that there are two implications. The fact that the algorithms used to represent crowds are still dense Dependency is the most significant issue that has been identified with the approaches now in use. Second, the methodologies of crowd modeling operate based on several rigorous and constrictive assumptions. This poll does not provide an in-depth investigation of crowd behavior in real-time events. Topics that cannot be Some of the topics discussed include high social force models, static crowd analysis, group profiling, and crowd behavior analysis. This study offers a comprehensive evaluation of the state-of-the-art methodologies in crowd behavior analysis from the points of view of both physics and biology. In addition to this, it offers debates that contribute to a more comprehensive understanding of how physics and some biology studies might be combined in a CV. This work aims to provide a foundation for future research on physics and biology-based approaches to computer vision. This report does not mention the stationary crowd or deep learning networks, which are both potential solutions to problems caused by crowd disasters caused by computer vision algorithms [20].

In recent years, research on crowd behavior has been conducted in various smart city applications, most notably in smart buildings [21, 22]. There have been several different solutions suggested, ranging from technological solutions based on ubiquitous computing for optimum monitoring [23–26], energy consumption analysis [27], and the utilization of machine learning for crowd prediction [28] and activity detection [29, 30]. [26] provides a comprehensive analysis of all of these different approaches to the problem. Analysis of pedestrian behavior is one of the other categories of smart city applications that have gotten comparatively less attention from the academic community.

Researchers have already done a lot of research regarding crowd density, crowd counting, crowd behavior, and crowd density. But still, there are gaps in the existing research. Such as, we cannot tell whether the crowd is violent or non-violent based on digital image processing. We can easily count the number of people in the crowd. We are able to detect the crowd density. But we are working on the detection of things in the hands of the crowd, including sticks or weapons, so we can decide whether this is a violent crowd or not. Let’s visualize the latest research trends regarding mob gathering and their behavior.

Different crowd counting methods have been suggested in the past, including static photos [31], [32] as well as video. One of the most widely used methods is counting with detection [33]. To forecast humans while they are walking, the moving windows-like detection is trained by the local visual features scanned over an im- age. When there are prominently displayed pedestrians in the image, this strategy is quite useful. Well-trained algorithms are necessary to capture the low-level properties of pedestrians. In comparison to counting by detection, a regression- based counting [34] procedure effectively removes the low-level information. The population count is estimated via the direct mapping between the low-level characteristics of the crowd and their density. Counting with clustering is a different method. This belief is that each artifact has distinct patterns that may be aggregated to approximate the crowd size. People are categorized into three main categories by [25]: macro-scope, micro-scope, and meso-scope.

### 2.1.2 Crowd tracking

Human or object tracking in all crowds, according to [37], entails monitoring the behavior-based information of the objects in the continuous digital image sequences of a video. To monitor an item, we must first recover the object’s temporal dependencies and then locate it in the continuous sequences of the images in the movie. The primary goal of monitoring or tracking is to pinpoint the location of objects. People shift their positions in each frame in congested circumstances. Tracking involves these places and calculating the route’s direction. Any anomalous occurrence can be quickly noticed and responded to with the assistance of the trajectory. Individuals in the crowd have a definite roadmap to follow [37]. The given route is represented by the trajectories. It is clear that a majority of individuals are implementing a set of instructions, and no abnormalities have been identified. [38] introduced the tracking technique that leverages a Correlated Topics Model (CTM) to follow people in unorganized populations, extracting the low-level motion data and mapping them across crowd behavior patterns. The advantages of employing the CTM are that no object detection is required, as well as low-level mob traits can be simply translated into various deviant behaviors. The research for this study depends on a combination of crowd tracking and behavioral research. [36] Work focused on tracking individual people inside the dense-crowd scenario while taking into account local as well as global dynamics at the same time. Personal behaviors are influenced by the scene’s geography, such as escape points, fences, and obstacles, as well as the motion of the other people. The fundamental concept is based on three evacuated dynamical designs: Stationary Floors Field (SFF), Dynamical Floors Field (DFF), and Boundary Floors Field (BFF), which are used to trace the influence of a scene as well as other passengers on many individuals’ behavior. [37] suggests an approach that uses neighboring movement concordance to identify as well as track people.

The mobility of many other people in the vicinity has an impact on the individual’s movement. According to [38], the general movement of a crowd is determined by the motion-based behavior of the individuals inside a crowd. They developed the Bayesian framework model for tracking visuals in a crowd, which uses the space-time concept. [39] presented the real-time technique, AdaPT, to determine individual trajectories in the crowd scenario. A Hidden-Markov classifier is con- structed on the local spatial-temporal characteristics of the individuals to reflect general crowded movement. It’s called the AdaPT because an algorithm trained about the parameters of every pedestrian in pictures starts some frames as well as continuously adjusts to calculate the trajectories of walkers. This strategy, on the other hand, works effectively in crowds of less than ten people. Their method depends on adaptive average shifting as well as particle tracking. [40] employed combined and separate pedestrian groupings to select fundamental classes, comprising merely two goals. For various-target surveillance, this grouping data is incorporated with just about any fundamental affinity structure. The clustering architecture of the tracks was encoded using a grouping network methodology.

### 2.1.3 Crowd Abnormality Detection

Another essential part of crowd analysis in video surveillance is abnormality detection inside the crowds [41]. In the video, the anomaly is characterized by abnormal patterns that differ from the otherwise regular happenings [42]. The identification of anomalies is represented in the spatiotemporal framework. The spatial-temporal component of irregularity detection [43]. The different behavior of crowds is primarily studied at 2 levels: locally as well as global. Independent action is assessed at the local level, whereas aggregate population movements are considered at the global level. Rather than establishing a person’s behavior, a global level is used to represent class behavior. The inappropriate behaviors at an individual level are depicted in (UCSD 2019). Irregular behavior can be identified in one of two different ways: activity monitoring or irregular population movement. [44] looked at the optical-flow characteristics of crowds and used spectral clustering to select the most appropriate number of frames to represent common movement patterns. Utilizing interactions [45], they proposed a new technique for detecting abnormalities in crowd videos that relies on pedestrian interaction forces. The interpersonal forces model is used to determine a relationship between various interaction forces (SFM).

The interaction pressure among moving elements is estimated by using SFM after the grid of the particles is laid across an image. The matrix of a particle’s location is altered utilizing the averaged optical flow field, as well as the interaction intensity amongst elements is determined. [46] suggested a technique for detecting anomalies. They used local and global identifiers to partition video frames into three dimensions. The descriptor-based similarity-based metrics were applied to identify abrupt fluctuations in spatial-temporal properties of contiguous regions. The closeness of the nearest neighboring sample is used to detect anomalous occurrences in [47].

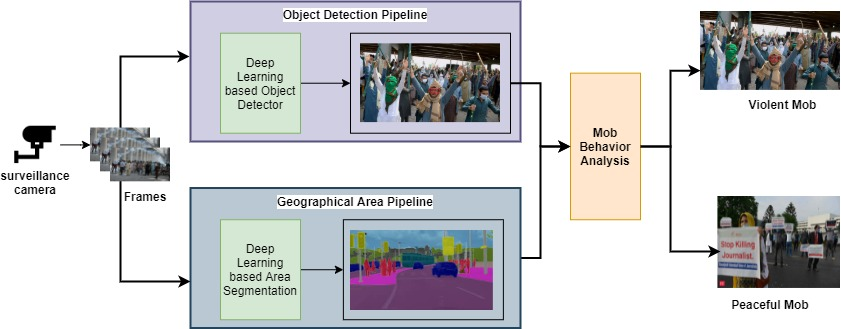
## 2.2 Research Gap

Previous work on mob analysis is quite limited, where the focus is mostly on finding the number of people and their density in a specified location. Each work had touched on a single aspect like crowd counting, crowd motion, and crowd behavior but lacked in determining the mob buildup and mob nature with prohibited objects.

# 3. Proposed Research Methodology

## 3.1 Proposed Framework

The proposed methodology consists of two main parts: the first part describes the object detection module, and the second part consists of the segmentation module. We have video frames from digital cameras that are mounted on the roadside. These image frames are passed through the object detection pipeline as well as the geographical area pipeline simultaneously. From this process, we have found crowd behavior analytics. In the object detection pipeline, we have a deep learning-based object detector (EfficientDet) that detects the objects that may be held by the crowd if it is a violent mob. Meanwhile, in the geographical area pipeline, we have a deep learning-based area segmentation (EfficientDet) of these frames to find out whether this crowd is on the road or not. This complete information about crowd-handling objects and roads is used to analyze the overall behavior of the crowd. At this stage, we can predict the mob’s behavior: either this is a peaceful crowd or they are violent people.



**Figure 1: Proposed Methodology Diagram**

## 3.2 Dataset

As we know, deep learning-based models are data hungry, and it requires data to train models. If the amount of data increased, the trained model would work accurately and perform in a better way. In our case, we have used EfficientNet for detection and segmentation purposes, so we need a dataset. We have searched a lot for the required dataset, but we did not find any publicly available dataset that fulfills our desired objectives. Then we used the JHU crowd++ dataset, which is publicly available but collected in a controlled environment. As in our case, we need to detect mob behavior in real time, so this publicly available dataset was used to generate results for different object detection algorithms. This dataset contains 4,372 images, but all the images were captured in a controlled environment. The resolution of these images is 1430x910. Secondly, we have collected our own custom dataset in an uncontrolled environment, which is much more robust. This self- created dataset contains 3,500 images with a 2024x1080 image resolution.



**Figure 2: JHU-CROWD Dataset Samples**

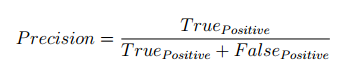


**Figure 3: Self Collected Dataset Samples**

# 4. Evaluation matrix

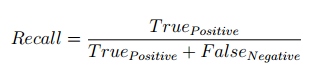
## 4.1 Precision

Precision, or the caliber of the successful predictions made with the algorithm, is indeed one measure of the algorithm’s performance. A Precision is calculated by dividing the complete numbers of the positive forecasts by the proportion of genuine positives (for example, the true positives and the false-positives). For instance, in the customer’s attrition algorithm, precision is a ratio of the total number of customers model properly anticipated to disconnect to the number of customers who actually did so are illustrated in Equation below



## 4.2 Recall

The number of positive examples is determined by dividing the total number of positive examples by the number of positive examples accurately categorized as ”positive”. The algorithm’s ability to recognize positive examples is measured by a recall. The greater the recall, the greater the number of positive examples found. This equation is shown under Equation



# References

[1] Saining Xie, Ross Girshick, Piotr Doll´ar, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1492–1500, 2017.

[2] Beibei Zhan, Dorothy N Monekosso, Paolo Remagnino, Sergio A Velastin,and Li-Qun Xu. Crowd analysis: a survey. Machine Vision and Applications, 19(5):345–357, 2008.

[3] Hina Uttam Keval. Effective design, configuration, and use of digital CCTV. PhD thesis, UCL (University College London), 2009.

[4] Nemanja Petrovic, Nebojsa Jojic, and Thomas S Huang. Adaptive video fast forward. Multimedia Tools and Applications, 26(3):327–344, 2005.

[5] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. Pattern recognition, 36(3):585–601, 2003.

[6] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3): 334–352, 2004.

[7] Julio Cezar Silveira Jacques Junior, Soraia Raupp Musse, and Claudio Rosito Jung. Crowd analysis using computer vision techniques. IEEE Signal Processing Magazine, 27(5):66–77, 2010.

[8] Joshua Candamo, Matthew Shreve, Dmitry B Goldgof, Deborah B Sapper, and Rangachar Kasturi. Understanding transit scenes: A survey on human behavior-recognition algorithms. IEEE transactions on intelligent transporta- tion systems, 11(1):206–224, 2009.

[9] NNA Sjarif, Siti Mariyam Shamsuddin, and SZ Hashim. Detection of abnor- mal behaviors in crowd scene: a review. Int. J. Advance. Soft Comput. Appl, 4(1):1–33, 2012.

[10] Angela A Sodemann, Matthew P Ross, and Brett J Borghetti. A review of anomaly detection in automated surveillance. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6):1257–1272, 2012.

[11] Alexandros Andr´e Chaaraoui, Pau ClimentP´erez, and Francisco Fl´orezRevuelta. A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Expert Systems with Applications, 39(12):10873– 10888, 2012.

[12] D Gowsikhaa, S Abirami, and Ramachandran Baskaran. Automated human behavior analysis from surveillance videos: a survey. Artificial Intelligence Review, 42(4):747–765, 2014.

[13] Myo Thida, Yoke Leng Yong, Pau Climent-P´erez, How-lung Eng, and Paolo Remagnino. A literature review on video analytics of crowded scenes. Intel- ligent multimedia surveillance, pages 17–36, 2013.

[14] Hyunjoon Jo, Kabir Chug, and Ricky J Sethi. A review of physics-based meth- ods for group and crowd analysis in computer vision. Journal of Postdoctoral Research, 1(1):4–7, 2013.

[15] Sarvesh Vishwakarma and Anupam Agrawal. A survey on activity recognition and behavior understanding in video surveillance. The Visual Computer, 29 (10):983–1009, 2013.

[16] Palwasha Afsar, Paulo Cortez, and Henrique Santos. Automatic visual de- tection of human behavior: A review from 2000 to 2014. Expert Systems with Applications, 42(20):6935–6956, 2015.

[17] Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan. Crowded scene analysis: A survey. IEEE transactions on circuits and systems for video technology, 25(3):367–386, 2014.

[18] M Sami Zitouni, Harish Bhaskar, J Dias, and Mohammed E Al-Mualla. Ad- vances and trends in visual crowd analysis: A systematic survey and evalua- tion of crowd modelling techniques. Neurocomputing, 186:139–159, 2016. [21] Ven Jyn Kok, Mei Kuan Lim, and Chee Seng Chan. Crowd behavior analysis: A review where physics meets biology. Neurocomputing, 177:342–362, 2016.

[19] Mohammad Mehedi Hassan, Md Golam Rabiul Alam, Md Zia Uddin, Shamsul Huda, Ahmad Almogren, and Giancarlo Fortino. Human emotion recognition using deep belief network architecture. Information Fusion, 51:10–18, 2019.

[20] Jiaqi Gong, Yu Huang, Philip I Chow, Karl Fua, Matthew S Gerber, Bethany A Teachman, and Laura E Barnes. Understanding behavioral dy- namics of social anxiety among college students through smartphone sensors. Information Fusion, 49:57–68, 2019.

[21] Noureddine Lasla, Messaoud Doudou, Djamel Djenouri, Abdelraouf Ouad- jaout, and Cherif Zizoua. Wireless energy efficient occupancy-monitoring system for smart buildings. Pervasive and Mobile Computing, 59:101037, 2019.

[22] Roufaida Laidi and Djamel Djenouri. Udeploy: User-driven learning for oc- cupancy sensors deployment in smart buildings. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (Per- Com Workshops), pages 209–214. IEEE, 2018.

[23] Dejun Yang, Satyajayant Misra, Xi Fang, Guoliang Xue, and Junshan Zhang. Two-tiered constrained relay node placement in wireless sensor networks: Computational complexity and efficient approximations. IEEE Transactions on Mobile Computing, 11(8):1399–1411, 2011.

[24] Marc Ringel, Roufaida Laidi, and Djamel Djenouri. Multiple benefits through smart home energy management solutions—a simulation-based case study of a single-family-house in algeria and germany. Energies, 12(8):1537, 2019. [28] Dustin Bales, Pablo A Tarazaga, Mary Kasarda, Dhruv Batra, Americo G Woolard, Jeffrey D Poston, and VVN S Malladi. Gender classification of walkers via underfloor accelerometer measurements. IEEE Internet of Things Journal, 3(6):1259–1266, 2016.

[25] HM Sajjad Hossain, Md Abdullah Al Hafiz Khan, and Nirmalya Roy. Active learning enabled activity recognition. Pervasive and Mobile Computing, 38: 312–330, 2017.

[26] Yi-Ting Chiang, Ching-Hu Lu, and Jane Yung-jen Hsu. A feature-based knowledge transfer framework for cross-environment activity recognition to- ward smart home applications. IEEE Transactions on Human-Machine Sys- tems, 47(3):310–322, 2017.

[27] Edward Toth and Sanjay Chawla. Group deviation detection methods: a survey. ACM Computing Surveys (CSUR), 51(4):1–38, 2018.

[28] Raghavendra Chalapathy, Edward Toth, and Sanjay Chawla. Group anomaly detection using deep generative models. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 173–189. Springer, 2018.

[29] Guanting Tang, Jian Pei, James Bailey, and Guozhu Dong. Mining multi- dimensional contextual outliers from categorical relational data. Intelligent Data Analysis, 19(5):1171–1192, 2015.

[30] Junli Li, Jifu Zhang, Ning Pang, and Xiao Qin. Weighted outlier detection of high-dimensional categorical data using feature grouping. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(11):4295–4308, 2018.

[31] Xujun Zhao, Jifu Zhang, Xiao Qin, Jianghui Cai, and Yang Ma. Parallel mining of contextual outlier using sparse subspace. Expert Systems with Ap- plications, 126:158–170, 2019.

[32] Meimei Gong and Yiming Shu. Real-time detection and motion recognition of human moving objects based on deep learning and multi-scale feature fusion in video. IEEE Access, 8:25811–25822, 2020.

[33] Yu Hao, Zhi-Jie Xu, Ying Liu, Jing Wang, and Jiu-Lun Fan. Effective crowd anomaly detection through spatio-temporal texture analysis. International Journal of Automation and Computing, 16(1):27–39, 2019.

[34] David Held, Sebastian Thrun, and Silvio Savarese. Learning to track at 100 fps with deep regression networks. In European conference on computer vision, pages 749–765. Springer, 2016.

[35] Xing Hu, Yingping Huang, Qianqian Duan, Wenyan Ci, Jian Dai, and Haima Yang. Abnormal event detection in crowded scenes using histogram of ori- ented contextual gradient descriptor. EURASIP Journal on Advances in Sig- nal Processing, 2018(1):1–15, 2018.

[36] Min Fu, Pei Xu, Xudong Li, Qihe Liu, Mao Ye, and Ce Zhu. Fast crowd den- sity estimation with convolutional neural networks. Engineering Applications of Artificial Intelligence, 43:81–88, 2015.

[37] Xianzhi Du, Mostafa El-Khamy, Jungwon Lee, and Larry Davis. Fused dnn: A deep neural network fusion approach to fast and robust pedestrian detection. In 2017 IEEE winter conference on applications of computer vision (WACV), pages 953–961. IEEE, 2017.

[38] Aniket Bera, Nico Galoppo, Dillon Sharlet, Adam Lake, and Dinesh Manocha. Adapt: real-time adaptive pedestrian tracking for crowded scenes. In 2014 IEEE International Conference on Robotics and Automation (ICRA), pages 1801–1808. IEEE, 2014.

[39] Gabriel J Brostow and Roberto Cipolla. Unsupervised bayesian detection of independent motion in crowds. In 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06), volume 1, pages 594– 601. IEEE, 2006.

[40] Chun-Yu Chen and Yu Shao. Crowd escape behavior detection and local- ization based on divergent centers. IEEE Sensors Journal, 15(4):2431–2439, 2014.

[41] Sheng Chen, Alan Fern, and Sinisa Todorovic. Person count localization in videos from noisy foreground and detections. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1364–1372, 2015.

[42] Haroon Idrees, Muhmmad Tayyab, Kishan Athrey, Dong Zhang, Somaya Al- Maadeed, Nasir Rajpoot, and Mubarak Shah. Composition loss for counting, density map estimation and localization in dense crowds. In Proceedings of the European conference on computer vision (ECCV), pages 532–546, 2018.

[43] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovit- skiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE conference on computer vi- sion and pattern recognition, pages 2462–2470, 2017.

[44] Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, and Ling Shao. Crowd counting and density estimation by trellis encoder-decoder networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 6133–6142, 2019.

[45] Di Kang, Zheng Ma, and Antoni B Chan. Beyond counting: comparisons of density maps for crowd analysis tasks—counting, detection, and tracking. IEEE Transactions on Circuits and Systems for Video Technology, 29(5): 1408–1422, 2018.

[46] W Luo, J Xing, X Zhang, X Zhao, and TK Kim. Multiple object tracking: a literature review. arxiv preprint arxiv: 14097618. 2014.

[47] Jong Hyun Kim, Hyung Gil Hong, and Kang Ryoung Park. Convolutional neural network-based human detection in nighttime images using visible light camera sensors. Sensors, 17(5):1065, 2017.