Ministry of Science and Higher Education of the Republic of Kazakhstan

Suleyman Demirel University

Mohamed Ibrahim

Image Based People Counting System for High Dense crowd

THESIS

Presented in Partial Fulfilment for the

Master of Technical Sciences Degree in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Associate Professor,PhD Cemil Turan**

Kaskelen 2023

Suleyman Demirel University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty

Associate Professor

PhD Zhamanov A.

____

06 ______ 06 ______ 2023

Topic of the thesis:

Image Based People Counting System for high Dense crowd

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science" SDU, 2021-2023

Head of Department ______ Assistant Professor, PhD Mukash Zh.

Academic Supervisor ______ Associate Professor, Cemil Turan.

Master student ______ Ibrahim M.

Kaskelen 2023

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Mohamed Ibrahim

2023

Acknowledgements

I would like to extend my deepest appreciation and gratitude to all those who helped finish my master's thesis. This research project would not have been possible without their support, direction, and inspiration. First and foremost, I want to express my deepest appreciation to Prof. Cemil Turan, who is my thesis supervisor, for his helpful advice, determination, and constant encouragement during the whole study process. Their knowledge, intelligent comments, and helpful critique were crucial in determining the course of this thesis. They pushed me to do my best work and were constantly there for me, and I owe them for that. I am grateful for the computer science professors in the computer scienenc Department at Suleyman Demirel University, whose comprehensive knowledge and instruction have given me a strong foundation in my field of competence. Their discussions, seminars, and lectures have broadened my understanding and piqued my interest in the subject. I want to express my appreciation to Suleyman Demirel University's personnel and leadership for creating a supportive academic environment for my research. They have proven incredibly helpful with their effective assistance, administrative support, and resource access. I would like to express my gratitude to my friends and coworkers for their continuous support, inspiration, and enlightening conversations along this trip. Their companionship, constructive evaluation, and idea-sharing have all greatly influenced my study. Last but not least, I want to thank my family from the bottom of my heart for their unending support, love, and faith in my skills. Their persistent assistance, understanding, and sacrifice have been valuable.

Dedication

I dedicate my master's thesis, "Image-Based People Counting system for high dense crowd" to all of the people who contributed significantly towards my academic and personal development and whose unwavering love, support, and encouragement made this achievement possible. I devote this thesis, foremost to my family. To my parents, who have been my pillars of support throughout my academic endeavors and who never stopped having faith in me. My achievement has been fueled by your selflessness, sound advice, and unchanged faith in my talents. This dissertation is proof of your constant love for me and the principles you have ingrained in me. In honor of my mother, who has supported me through the highs and lows of this academic journey, I dedicate this work to her. Your, steadfast assistance, compassion, and inspiration have served as my continual sources of inspiration. I appreciate your support and your being my source of confidence and motivation. I also dedicate my thesis to Zamzam and Qadan, my siblings. Your assistance and inspiration have been essential, both academically and emotionally. Your trust in my potential and your eternal faith in my skills have continually strengthened my will to succeed. I want to express my dedication to my thesis advisor, Cemil Turan. This thesis's direction and caliber have been greatly affected by your advice, guidance, and knowledge. I thank you for your unwavering assistance, tolerance, and Insightful comments have been crucial to my development as a scholar. I would like to thank the professors at Suleyman Demirel University, whose commitment to excellence and teaching has improved my academic career. My perspectives have been opened up by their knowledge, enthusiasm, and intellectual direction, and they have inspired me to strive for excellence within my field of study. I dedicate this thesis to my friends and coworkers who regularly offer moral encouragement and intellectual stimulation, and their friendship, stimulating conversations, and experiences we have shared have deepened and made my academic life even more delightful. Last but not least, I dedicate this thesis to everyone who has assisted me in enhancing myself, whether it be through relationships, mentorship, or shared knowledge. Your influence on my professional development has been immense, both personally and professionally, and I am thankful for the expertise acquired and the memories that were made.

I would like to send my deepest thanks and respect to everyone who has taken part in this wonderful excitement. I dedicate this thesis to every single one of you because it constitutes a testimony to our collective work.

Abstract

The main objective of crowd counting is to count people in crowd images accurately. Contemporary methodologies use multi column CNN designs to regress density maps of crowd images in order to operate the scale or perspective shifts which frequently appear in crowd images. Considering advances in technology, low-cost camera surveillance is already widely deployed. They are used for surveillance in important locations like buildings and businesses as well as in common areas like parks, schools, railways, and airports. Typically, the things in question are moving people or moving automobiles. Crowd analysis is essential in safety evaluation and other related domains. But most of the work was done by people up until this point. Automatic assessment of crowds is becoming more and more necessary when there is a lot of photographic equipment because manual approaches may become unreliable or expensive. Population counting is the first aspect of crowd evaluation that may be dealt with as a computer vision problem. It creates persons in the image using a data image.Such detection-based systems have the drawback that the performance of the detector is strongly impacted by the presence of people in congested areas or large crowds, which reduces the accuracy of the final estimation. People have suggested clustering the trajectories of monitored visual characteristics in order to count crowds in videos. Various columns' reception areas respond differently to different sized people's bodies.our main objective is to make a system that can count high dense crowd accurately by using images of the people.

Аңдатпа

Көпшілікті санаудың негізгі мақсаты - тобыр бейнелеріндегі адамдарды дәл санау. Қазіргі заманғы әдістемелер масштабты немесе перспективалық жылжуларды басқару үшін топтық кескіндердің тығыздық карталарын регрессиялау үшін көп бағаналы CNN дизайнын пайдаланады. тобырдың суреттерінде жиі пайда болады. Технологияның жетістіктерін ескере отырып, арзан камералық бақылау қазірдің өзінде кеңінен қолданылады. Олар ғимараттар мен кәсіпорындар сияқты маңызды орындарда, сондай-ақ саябақтар, мектептер, темір жолдар және әуежайлар сияқты жалпы аумақтарда бақылау үшін пайдаланылады. Әдетте, бұл адамдар қозғалатын немесе қозғалатын автомобильдер. Қауіпсіздікті бағалауда және басқа қатысты домендерде топты талдау маңызды. Бірақ осы уақытқа дейін жұмыстың көп бөлігін адамдар атқарды. Көпшілікті автоматты түрде бағалау фотографиялық жабдық көп болған кезде қажет болып барады, өйткені қолмен тәсілдер сенімсіз немесе қымбат болуы мүмкін. Популяцияны санау - компьютерлік көру мәселесі ретінде қарастырылуы мүмкін топты бағалаудың бірінші аспектісі. Ол деректер кескінін пайдаланып кескінде адамдарды жасайды. Мұндай анықтауға негізделген жүйелердің кемшілігі бар, бұл детектордың жұмысына кептеліс аумақтарындағы немесе көп адамдар жиналатын адамдардың болуы қатты әсер етеді, бұл соңғы бағалаудың дәлдігін төмендетеді. Адамдар бейнелердегі топтарды санау үшін бақыланатын визуалды сипаттамалардың траекторияларын кластерлеуді ұсынды. Әртүрлі колонналардың қабылдау аймақтары әртүрлі өлшемдегі адамдардың денесіне әртүрлі жауап береді. Біздің басты мақсатымыз - бұл адамдардың суреттерін пайдалану арқылы жоғары тығыз топты дәл санай алатын жүйе.

Аннотация

Основная цель подсчета толпы - точно подсчитать людей на изображениях толпы. Современные методологии используют многоколоночные конструкции CNN для регрессии карт плотности изображений толпы, чтобы управлять масштабом или сдвигами перспективы. которые часто появляются на изображениях толпы. Учитывая достижения в области технологий, недорогие камеры видеонаблюдения уже широко используются. Они используются для наблюдения в важных местах, таких как здания и предприятия, а также в местах общего пользования, таких как парки, школы, железные дороги и аэропорты. Как правило, речь идет о передвижении людей или передвижении автомобилей. Анализ толпы необходим для оценки безопасности и других смежных областей. Но большую часть работы до этого момента выполняли люди. Автоматическая оценка скопления людей становится все более необходимой при большом количестве фотооборудования, поскольку ручные подходы могут стать ненадежными или дорогими. Подсчет населения — это первый аспект оценки толпы, который можно рассматривать как проблему компьютерного зрения. Он создает людей на изображении с использованием изображения данных. Недостаток таких систем на основе обнаружения заключается в том, что на производительность детектора сильно влияет присутствие людей в перегруженных местах или больших скоплениях людей, что снижает точность окончательной оценки. Люди предложили группировать траектории отслеживаемых визуальных характеристик, чтобы подсчитывать толпы на видео. Приемные зоны различных колонн по-разному реагируют на тела людей разного размера. система, которая может точно подсчитывать высокую плотность скопления людей, используя изображения людей

Abbreviations

RRR ridge regression

MORR multi-output ridge regression

GPR Gaus- sian processes regression

CNN convolutional neural network

GLCM Gray Level Co-occurrence Matrix

SURF Speeded Up Robust Features

HOG histogram of oriented gradients

KLTracker Kanade–Lucas–Tomasi

Table of Contents

Chapter 1

Background and motivations

1.1 Introduction

Crowd counting, public space design, and pedestrian profiling are only a few of the various applications for crowd counts throughout public areas. For counting the crowds on a railway platform, for example, assuming an overall count across the whole scene is adequate. It is additionally important to figure out the number of people in different spatial places in more complex scenarios. For example, when performing a crowd count in a mall, it is essential to comprehend not only the overall number of individuals there but also how they are scattered, or more specifically, which shop is busier. Crowd oversight, public space design, and pedestrian profiling were merely a few of the various applications for population counting in public areas.

Despite the rise in technology, inexpensive surveillance with cameras is now frequently used. They serve for surveillance in public places like parks, schools, stations, and airports, as well as in major locations like structures and companies. Humans or vehicles in motion are typically the objects under consideration. In safety evaluation and other related fields, crowd analysis is crucial. But up until now, a large portion of all the tasks were done by people. When there is a lot of photographic equipment, using manual methods may turn unreliable or costly, making automatic assessment of crowds increasingly needed. The initial component of crowd assessment that can be handled as a computer vision issue is population counting. It takes an image in the data and generates people in a particular image. This number may be utilized to assess the efficacy of an occasion or avoid a crowded scenario. Social prejudice makes physical crowd counting hard and inaccurate. Crowd counting is fairly adaptable across multiple fields. Effective crowd counting approaches can be extended to other fields, including cell and automobile counting. Crowd counting has multiple practical applications, such as security monitoring, preventing catastrophes, designing public areas, collecting and evaluating intelligence, simulated settings, and forensics searches. Crowd counting can be used in security surveillance to analyze crowd behavior in video footage. Crowd counting can be used in emergencies to find overcrowding. Building public spaces with population flow and the results of enumeration can be employed to improve the design of future buildings. Crowd counting can be used in data collection and analysis to gauge the public's interest. In overcrowded environments, counting crowds strategies for forensic search may enhance conventional methods of recognizing faces and identification.

On the other side, places, including cinemas, sports venues, events, speeches about politics, and gatherings, feature dense crowds. one picture only. In situations where the scene is in itself unconstrained, a method of "public count" is used for estimating the number of free goods (such as people, cars, cells, etc.). in a scenario. Because of its vital uses in fields such as customer behavior, traffic control, public safety, and cell counts, crowd counting is currently attracting a lot of interest. In understanding the significance of carry forecasting, many techniques have been employed, especially the use of machine learning, which is more efficient for a variety of uses, including testing hypotheses, image categorization, and several kinds of co-time series. Crowd counting is used to determine how many people appear in images and movies. It is an important topic with numerous practical uses, like video surveillance (such as complaint handling in a particular location or finding densely packed audiences), security management (such as calculating the number of individuals who enter or leave certain regions), web multimedia (such as the size of crowd estimates for tweet image launching within a crowded scenic spot), as well as more. It's an important subject in many real-world uses, such as surveillance footage (such as complaint management for a specific region or identifying crowded audience participants), security monitoring (for instance, tallying the number of people who enter and leave only a few zones), and the internet (for instance, the size of the crowd forecasting over tweet shootings against a crowded scenic background); a demographic census is an essential endeavor that could truly utilize population statistics. There remain numerous real-world uses for load prediction, such as forensic search, disaster management, the creation of broad areas, collecting and analyzing information, and the use of simulated worlds. Monitoring methods like crowd counting can be used to analyze human behavior in images. Crowding can be detected by crowd qualifiers in the prevention of disasters. The results of studies on crowd movements or numbers of people can be utilized to improve designs for buildings in public areas. Crowd counting might be used for gathering and analyzing information to determine the public's interest. Crowd-counting methods utilized in forensic searches might improve current techniques for identifying faces.

There are currently three types of people counting methods in use: counting by detection, counting by clustering,and counting by regression. In counting by detection [1].the total number of people is determined through the recognition of human occurrences. The detection technique can be lengthy because it entails comprehensive picture field scanning using a trained detector with different scales[2]. Measuring the number of people in dense crowds poses challenges. In between the identification problem and the regression problem comes the intriguing problem of the population count. Regression-based ideas are suggested by acquiring the mapping between the global and localized images and their corresponding crowd numbers to address the counting of crowds as a regression problem. Models of regression typically perform effectively when the head dimensions are small, but if the overall head size gets large, their accuracy degrades. Nevertheless, spatial information, such as the variation in density distribution across the image or the association of nearby localized patches, is ignored or not properly used. The significant use of such information, therefore, makes sense for dense map-based methods[3].They develop a model that produces a density. A significant use of this data is therefore suggested using density map-based techniques. They build a model that produces a map of density showing the total number of individuals at the pixel level. But there is also the issue of inconsistent spread in density maps. Some techniques attempt to divide the image into smaller areas despite assuming that each area's density is constant. These approaches could still be vulnerable to boundary loss buildup and how local regions are collected. In addition, existing crowd-counting datasets' basic facts are a bit uncertain. Because annotating data is costly, the available data are limited in terms of either their resolution or the number of participants. A crowd is thought to be formed up of separate things for the count by clustering, every one of which has unique but coherent movement characteristics as is in the below (see Figure 1.1) these methods might still be susceptible to border loss accumulation and the way local areas are gathered. Basic facts in the available crowd-counting databases are also a little hazy. The accessible data are constrained in terms of either their resolution or the quantity of participants because annotating data is expensive to perform.

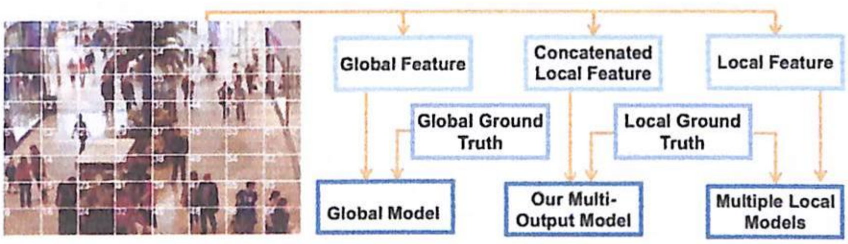

Figure 1.1: Flow chart that shows how our multioutput model and regression analyses are used for detecting crowd counting

which can be clustered to approximate the number of people. Only films with an adequate rate of frames enable the accurate recovery of all motion data utilizing a method that utilizes movement clusters. Focal point monitoring, or intentional segmentation of objects, is employed in the count by recognition and classification techniques. These are not suitable for occupied circumstances that have a dense background and frequently occurring inter-object shadowing. Rather than merely separating and tracking people, the counting by regression approach wants an immediate relationship between modest variables and the overall number of people. This approach is, therefore, more effective financially and is, therefore, more appropriate for areas that are crowded. There are two types of counting by-regression techniques now in use: global methods and local methods (see Figure 1). A single regress function relating the overall number of people in the image and the total number of image features gathered across the entire image space can be learned by the global method[4]. That is, geographic data disappears while estimating global characteristics; a representation like this automatically presumes that a feature should have the same weight no matter where in the scene it is retrieved. In several circumstances found in real life, this presumption is usually false. Due to the density, scene arrangement, or crowd autonomy brought about by basic human conversations, boundaries, and regulations, crowded structures are especially able to shift spatially[5].As a consequence, various characteristics could prove more precise and useful in crowd counting in particular spatial situations. Furthermore, the global model of regression cannot offer information regarding geographically restricted population counts, as is needed in certain situations[6]. By separating the image space into cell portions, every one of which is represented using a distinct extrapolation operation, local simulations [7]. It also attempts to loosen the global premise to a certain degree to deal with the drawbacks of a global strategy. To account for camera geometrical errors, the zones could consist of either regular-sized cells or a varied resolution determined by the scene perspective. A global count can be derived by combining the cell-level numbers after local quantities have been computed in each region. Approach the challenge of anticipating image densities, whose integral over each image region provides the count of objects inside that region to an extremity, by modeling the crowd densities at every pixel in the image. Localized models typically try to assess characteristics differentially by local crowd structure, in contrast with global methods[8]. As they encourage localized crowd counting. Nevertheless, since various regression models have to be learned, and there can be quite a few of them, the scalability of current local approaches is an issue. The reality that no data is transmitted between geographically localized locations to provide a more context-aware feature selection for more precise crowd counting is another fundamental flaw of existing regional models. Because of the cluttered backdrop as well as significant inter-object obstructions, minimal imaging characteristics are frequently extremely unclear in everyday situations. Considering common characteristics and traits among diverse local regions would therefore assist in determining the number of people[9]. Regarding precise and trustworthy crowd counting, both are missing across all current techniques; local feature significance extraction and data return throughout fields constitute two key elements. For us to accomplish this, we offer a single multi-output model built around regression using ridges that take dependent regional variables from small spatial areas as input and produce multi-dimensionally arranged outputs from population numbers in distinct geographic areas (see Figure 1). Our approach relaxes the one-to-one modeling needed in contrast to global regression techniques by acquiring regionally local regression parameters simultaneously in one model across all the distinct cell regions within the image[10]. Models can localize trait relevance. Our one model, as opposed to current methods for creating a few local models of regression, undergoes training through collaborative optimization to enforce mutual dependence across cell regions. As a result, data from all regional areas can be swapped to generate an occupancy forecast that becomes more precise[11]. The main outcomes of this work are listed below: This has been the first study to achieve accurate crowd counting through exchanging visual information among regionally localized regions in an area and mining specific characteristics for importance. For localized crowd counting, this is achieved by taking into consideration just one multi-output ridge correlation approach that provides benefits over either current worldwide methods when it comes to providing local estimations or existing local methods in terms of being easier to scale. For crowd analysis, we provided a new public scene data set containing over 60,000 pedestrian instances. It includes the most realistic and challenging setting of a stuffed situation in an open space that we know of, resulting in the biggest data set so far.

1.2 Aims and Objectives

As a part of our work, we set ourselves the goal of implementing a people-counting system for highly dense crowds. This can help with many problems in the world. For instance, it can help management with safety. Unusual behavior Detection Political significance of rallies or protests. An image-based system for counting in highly dense crowds was developed and researched. We set ourselves the task of creating an image-based system that can detect highly dense crowd areas that could impact many areas. Properly monitoring people in crowd images is the major goal of crowd counting. To handle the scale or perspective alterations that typically arise in crowd photos, modern approaches use multi-column CNN designs to regress density maps of crowd images. The receiving rooms of different columns react differently to people of various sizes. The main goal is to create a system that can accurately count large, dense crowds of people using photographs of the population. The aims and objectives of crowd counting can be categorized as follows: Security as Security: Providing that everyone is secure constitutes several of the primary objectives for the human count in areas with high populations. Officials can track the level of crowding to take measures to prevent crowding, stampedes, or other potentially hazardous situations by carefully measuring those who exist in a crowded area. Crowd administration: Counting people aids in efficient crowd management. Event organizers or security personnel may utilize methods that maximize crowd flexibility, reduce bottlenecks, and ensure a smooth flow between individuals within the space by investigating crowd demographics and flow patterns. Resource Allocation: Accurate crowd-counting aids in resource allocation. The number of participants can be calculated by the organizers of the event, who can then allocate resources like furniture, drinks, and amenities accordingly. This ensures the facilities and infrastructure can meet the requirements of those in attendance. The term "real-occupant tracking" can be obtained through individual counts in places and organizations with a lot of visitors, such as arenas, malls, or transportation centers. Utilizing such information, managers of facilities can regulate entry to particular domains, track entry and exit points, and maintain a pleasant and secure environment for all guests to use. Disaster Planning: The most important aspect of preparing for an emergency is people. Emergency workers can plan and carry out evacuations in high-density crowds by having a complete list of the people there. This assures that every person is safe regardless of a fire, an earthquake, or a terrorist threat. Crowd analytics: Material collected through person counts may be employed to assess the crowd's actions and trends. By analyzing the crowd dynamics, event planners and advertisers may choose properly whether it pertains to money allocations, spectator involvement strategies, or targeted advertisements. Crowd tracking can help manage traffic in crowded urban areas like transportation hubs. Authorities may alter signals, direct cars, and offer additional transit services by keeping tabs on how numerous individuals arrive or exit specific parts to try to lessen congestion. Utilizing intelligence from business and advertising Data on person counts can be applied to advertising and company statistical analysis. Shops may employ footfall traffic information to comprehend consumer behavior, plan out their stores, and assess the success of their promotional efforts. Making data-driven business choices while improving customer experiences may profit from this knowledge. Medical Services with Social Distance: Those who have worries about the health of everyone When there are outbreaks or epidemics of diseases, counting might assist and facilitate the enforcement of communal isolation measures. Governments may mitigate the possibility of spreading viruses and enforce conformity to safety standards by monitoring the number of people and implementing occupation constraints. Management utilizing prediction modeling: For planning and forecasting, long-term population-qualifying statistics can be used. Urban developers, planners for events, or transportation officials may predict future crowd sizes, prepare the creation of infrastructure, boost transportation services, or enhance the overall development of cities by analyzing past behavior. Crowd control: Counting people makes it easier to manage large crowds. By looking at crowd demographics and flow patterns, event planners or security professionals can employ strategies that increase crowd flexibility, minimize bottlenecks, and maintain a fluid flow amongst people inside the venue. Allocating resources is made easier by accurate crowd counting. The event planners can estimate the number of attendees and then distribute supplies like furniture, beverages, and amenities in accordance with that amount. This guarantees that the infrastructure and amenities can satisfy the needs of those in attendance. Individual counts at busy locations and establishments, such as arenas, shopping malls, or transportation hubs, can provide information on "real-occupant tracking." Using this data, facility managers can control access to specific domains, monitor Facilities can control access to specific domains, monitor entry and exit points, and uphold a comfortable and safe environment for all visitors to use. Planning for disasters: People are the most crucial component of emergency preparedness. When there are dense crowds, emergency personnel can organize and execute evacuations by having a complete list of everyone present. This guarantees everyone's safety regardless of a fire, an earthquake, or a danger from terrorism. Audience analysis To evaluate the crowd's behavior and patterns, data from person counts may be used. Event organizers and advertisers can make informed decisions about budget allocations, tactics for involving spectators, and targeted advertising by studying crowd dynamics. In dense metropolitan areas, crowd tracking can help manage traffic. like hubs for transit. By tracking how many people enter or leave particular areas, authorities can adjust signals, manage traffic, and provide more transit options in an effort to reduce congestion. using market and advertising intelligence Statistics for businesses and advertising can both benefit from person-count data. Shops can use foot traffic data to understand consumer behavior, layout their spaces, and evaluate the effectiveness of their marketing campaigns. This knowledge may be used to make data-driven business decisions while enhancing client experiences. Healthcare with Social Distinction: For those that are concerned about everyone's health when there are disease outbreaks or epidemics, counting could help facilitate the implementation of measures for societal isolation. Governments could lessen the likelihood of spreading By keeping track of the population and enforcing occupation restrictions, one can prevent the spread of diseases and ensure compliance with safety standards. Management that makes use of prediction modeling Long-term population-qualifying statistics can be utilized for planning and forecasting. By studying historical behavior, urban developers, event planners, or transportation officials can forecast future crowd numbers, plan the construction of infrastructure, improve transit services, or advance the general development of communities.

By keeping track of the population and enforcing occupation restrictions, one can prevent the spread of diseases and ensure compliance with safety standards. Management that makes use of prediction modeling Long-term population-qualifying statistics can be utilized for planning and forecasting. By studying historical behavior, urban developers, event planners, or transportation officials can forecast future crowd numbers, plan the construction of infrastructure, improve transit services, or advance the general development of communities.

maintain clear entry and departure points and a welcoming, secure environment for all visitors to use. Planning for disasters: People are the most crucial component of emergency preparedness. When there are dense crowds, emergency personnel can organize and execute evacuations by having a complete list of everyone present. This guarantees everyone's safety regardless of a fire, an earthquake, or a danger from terrorism. Crowd analytics: Information gathered through person counts may be used to evaluate the trends and behaviors of the crowd. Event organizers and advertisers can make informed decisions about budget allocations, tactics for involving spectators, and targeted advertising by studying crowd dynamics. With congested metropolitan locations like transit hubs, crowd tracking can aid with traffic management. Authorities could change or direct traffic, provide additional transit services, and keep track of how many people enter or leave particular areas in an effort to reduce congestion. using market and advertising intelligence Statistics for businesses and advertising can both benefit from person-count data. Shops can use foot traffic data to understand consumer behavior, layout their spaces, and evaluate the effectiveness of their marketing campaigns. This knowledge may be used to make data-driven business decisions while enhancing client experiences. Medical Services with Social Distancing: People who are concerned about everyone's health When there are disease outbreaks or epidemics, counting could help facilitate the implementation of measures for societal isolation. Governments can prevent the spread of diseases and ensure adherence to safety regulations by monitoring the population and putting occupation restrictions in place. Management that makes use of prediction modeling Long-term population-qualifying statistics can be utilized for planning and forecasting. By studying historical behavior, urban developers, event planners, or transportation officials can forecast future crowd numbers, plan the construction of infrastructure, improve transit services, or advance the general development of communities. Counting people makes it simpler to handle large crowds. Event organizers or security experts can use tactics to boost crowd adaptability, reduce bottlenecks, and maintain a fluid flow among individuals inside the venue by taking a look at crowd demographics and flow patterns. Accurate crowd counts facilitate resource allocation. The organizers of the event can determine the approximate number of guests before distributing supplies like furniture, refreshments, and amenities in keeping with that number. This ensures that the facilities and amenities will be adequate to meet attendees' needs. Individual counts at crowded places, like arenas, malls, or transportation hubs, can reveal data on "real-occupant tracking."

Using this information, facility managers can manage visitor access to particular domains, keep an eye on entry and departure points, and maintain a welcoming and secure environment for all users. preparing for emergencies: The most important aspect of emergency planning is people. Emergency workers can plan and carry out evacuations when there are large crowds by having a thorough list of everyone present. This ensures everyone's security regardless of the threat of terrorism, fire, or earthquake. audience research To Data from person counts may be utilized to analyze the crowd's patterns of activity. By analyzing crowd dynamics, event planners and advertisers may make wise decisions about budget allocations, strategies for engaging spectators, and targeted advertising. Crowd tracking can assist in traffic management in crowded urban areas. like transportation hubs. Authorities may regulate traffic, change traffic signals, and add more transit alternatives by keeping track of how many people enter and exit specific locations. Using advertising and market research Data on the number of people can be used to inform both company statistics and advertising. Data on foot traffic can be used by retailers to plan their spaces, study consumer behavior, and assess the success of their marketing initiatives. Making business judgments based on data is possible with this information is possible.improving customer experiences Those who care about everyone's health are those who practice healthcare with social distinction. Counting could aid and facilitate the implementation of measures for societal isolation during disease outbreaks or epidemics. Governments may reduce the possibility of spreading One can stop the spread of diseases and guarantee adherence to safety standards by monitoring the population and implementing occupation limitations. the application of predictive modeling in management For planning and forecasting, long-term population-qualifying statistics can be used. Urban planners, event organizers, or transportation authorities can anticipate future crowd sizes, plan the building of infrastructure, enhance transit services, or advance the general development of communities by looking at historical behavior. Counting individuals can help with effective crowd control. By looking at crowd demographics and flow patterns, event planners or security professionals can employ strategies that increase crowd flexibility, minimize bottlenecks, and maintain a fluid flow amongst people inside the venue. Allocating resources is made easier by accurate crowd counting. The event planners can estimate the number of attendees and then distribute supplies like furniture, beverages, and amenities in accordance with that amount. This guarantees that the infrastructure and amenities can satisfy the needs of those in attendance. Individual counts at busy locations and establishments, such as arenas, shopping malls, or transportation hubs, can provide information on "real-occupant tracking." Using this data, facility managers can control access to specific domains, monitor and maintain clear entry and departure points, and create a welcoming, secure environment for all visitors to use. Planning for disasters: People are the most crucial component of emergency preparedness. When there are dense crowds, emergency personnel can organize and execute evacuations by having a complete list of everyone present. This guarantees everyone's safety regardless of a fire, an earthquake, or a danger from terrorism. Crowd analytics: Information gathered through person counts may be used to evaluate the trends and behaviors of the crowd. Event organizers and advertisers can make informed decisions about budget allocations, tactics for involving spectators, and targeted advertising by studying crowd dynamics. With congested metropolitan locations like transit hubs, crowd tracking can aid with traffic management. Authorities could change or direct traffic, provide additional transit services, and keep track of how many people enter or leave particular areas in an effort to reduce congestion. using market and advertising intelligence Statistics for businesses and advertising can both benefit from person-count data. Shops can use foot traffic data to understand consumer behavior, layout their spaces, and evaluate the effectiveness of their marketing campaigns. This knowledge may be used to make data-driven business decisions while enhancing client experiences. Medical Services with Social Distancing: People who are concerned about everyone's health. When there are disease outbreaks or epidemics, counting could help facilitate the implementation of measures for societal isolation. Governments can prevent the spread of diseases and ensure adherence to safety regulations by monitoring the population and putting occupation restrictions in place. Management that makes use of prediction modeling Long-term population-qualifying statistics can be utilized for planning and forecasting. By studying historical behavior, urban developers, event planners, or transportation officials can forecast future crowd numbers, plan the construction of infrastructure, improve transit services, or advance the general development of communities.

1.3 Thesis Outline

The beginning chapter is the introduction. The general comprehension and essential elements of the image-based people counting system for a highly dense crowd were covered in the introduction. the pros and cons of counting dense crowds,as well as its main key elements. This chapter also outlines my goals and our rationale for choosing this topic to write about. A review of the literature is in the second chapter. This chapter included writings from researchers that were in some way linked to our topic. Types of people-counting methods are covered in the third chapter. This chapter describes many types of people-counting methods, how they operate, and their benefits and drawbacks. Additionally, articles comparing the fundamental methods used by image-based people counting systems for highly dense crowds were implemented. The implementation of the people-counting system is covered in the fourth chapter. And finally, we conclude the discussion in the chapter on improvement in the future.

Chapter 2

Literature Review

2.0.1 Review of existing solutions

There are only two main categories of crowd-counting methods: comprehensive and localized. In holistic approaches that use global image characteristics for assessing every frame in a video series, the space of features and the size of the crowd estimate are mapped using a classification and regression model. Local methods, on the other hand, make use of particular visual cues to locate, count, or measure pedestrians inside a given area of a picture. In this instance, the group's size is defined as its total membership. Many algorithms have been proposed in the literature for crowd counting. Earlier methods adopted a detection-style framework that scans a detector over two consecutive frames of a video sequence to estimate the number of pedestrians based on appearance and motion features. have used a similar detection-based framework for pedestrian counting. In detection-based crowd-counting methods, people typically assume a crowd is composed of individual entities that can be detected by some given detectors. The limitation of such detection-based methods is that the occlusion of people in a clustered environment or a very dense crowd significantly affects the performance of the detector, hence the final estimation accuracy. In counting crowds in videos, people have and probabilistically grouped them into clusters representing independently moving entities. However, such tracking-based methods do not work for estimating crowds from individual still images. Arguably the most extensively used method for crowd counting is feature-based regression (see. The main steps of this kind of method are segmenting the foreground and extracting various features from the foreground, such as the area of the crowd mask edge count or texture features, and utilizing a regression function to estimate the crowd counts. Linear or piecewise linear functions are relatively simple models and yield decent performance. Other more advanced and effective methods are ridge regression, Gaussian process regression, and neural networks. multiple sources of information to compute an estimate of the number of individuals present in an extremely dense crowd visible in a single image. In that work, a dataset of fifty crowd images annotated by humans is introduced. has followed the work and estimated counts by fusing information from multiple sources, namely, interest point Fourier analysis, wavelet decomposition, GLCM features, and low-confidence head detections. has utilized the features extracted from a pre-trained CNN to train a support vector machine that subsequently generates counts for still images. Recently, CNN has proposed a CNN-based method to count crowds in different scenes. They first pre-trained a network for certain scenes. When a test image from a new scene is given, they choose similar training data to fine-tune the pre-trained network based on the prospective information and similarity in the density map. Their method demonstrates good performance on most existing datasets. But their method requires perspective maps both on training scenes and the test scene. Unfortunately, in many practical applications of crowd counting, perspective maps are not readily available, which limits the applicability of such methods.

A compromise technique that uses length histograms based on local segments to convey the information fully has been proposed[12]. The holistic techniques are addressed, the subsequent approaches are discussed and the local approaches are examined.

2.1 Holistic methods

Holistic crowd counting approaches use global visual attributes to estimate the size of a crowd. These techniques are referred to as "map-ping-based" methods since they map exactly throughout the space of features with the crowd number estimate. While neural networks linear regression and the Gaussian procedure regression [13] all been used in categorization and regression techniques, they have all used characteristics such as patterns [14], the pixels that indicate the foreground of an image as well as distinctive edges[15]. Textural approaches rely on the proposition that crowded populations have coarse textures and sparsely populated crowds have subtle patterns. These approaches categorize the crowd densities using a four- or five-point system rather than simply determining the number of people. For estimating crowd density, there are suggestions using statistics centered on the gray level combination matrix (GLCM). The Minkowski fractal dimension was previously postulated. As for the suggested Translation Invariant Orthonormal Chebyshev Moments, it suggested using the 2D Discrete Wavelet Transform as the basis for extracting texture information. Because afternoon data had fewer variations in illumination than early data, their evaluation discovered that texture attributes performed well with an afternoon sample dataset. Because of these shifts in radiation throughout a period, efficiency suffered when both afternoon and morning specimens were merged to generate an expanded mixed set. This highlights the basic constraint of texture characteristics: the necessity for continuous retraining because their susceptibility to backdrop change renders them ineffective in the real world. Foreground pixels and edges are parameters that are used in other comprehensive crowd-counting techniques. While each of these features can be observed in areas of enticement, they are integrated on an overall basis. With the purpose of "detecting the bodies (i.e., legs and arms)," several algorithms have currently tried distinguishing the primary focus using modeling of the background methods. outlines the rationale for mainly those approaches. A human observer has no difficulty at all telling an extremely densely populated region from the background[16]. Following the theory, the had no difficulty at all distinguishing extremely densely populated regions from the background. The ratio of "crowd area" to "background region" is thought to serve as an accurate measure of the number of people in the brain of a person since it shows how trained it is. If the image's pixels that belong to the population were to be separated from those that relate to the scenery, the idea might be effectively used for the computerized estimation of density. In a study conducted, the connection between the number of pixels in the foreground and the total number of people in the image was roughly straight, similar to how it was for border pixels[17]. Together with employing border and foreground pixel counts, this also puts off an instant learning method for feedback neural network models. Each subregion corresponding to the image's background pixel percentage was computed, and the outcomes were applied to create a vector of features that served as an input to a network of neural networks for a regression[18]. The aforementioned techniques are accurate for interior crowd assessment over a short period of time. These approaches, however, utilized a static backdrop model, making the structure susceptible to shifts in lighting over a longer period of time. a combination of suddenly or progressively In recent uses for population counting, adaptable background models have been used since they're resilient to these alterations. In situations that used a slightly elevated angle of view, where the consequences of perspective weren't immediately apparent and the assessments were predicated[19]. Because distant things appear smaller and hence give fewer pixels to a foreground mask when viewpoint distortion is significant, the overall number of pixels in the foreground is not going to be an accurate indicator of overcrowding. Utilizing an exponential function of the column and row coordinates, the usage of quasi-calibration is presented, which is derived from the "relative variation in size of the projected height proportional widths for a rigid object as the object translates in depth"[20]. Utilizing the quasi-calibration that requires giving each of the pixels a weight that accounts for the effects of viewpoint, a "density map" is generated. It was discovered that severe overcrowding above a threshold might be identified using the weighted total of images of a foreground mask[21]. Similar methods were employed using a map of density to generate a weighted background pixel count and extrapolation using neural networks to figure out the size of a crowd[22]. There is also a comprehensive To be able to account for both occlusion and additional non-linearities like segmentation mistakes, the system extracted an enormous amount of information from each image. The dynamic texture served as a basis for segmentation, which produced two logical foreground masks: one for movement in every direction. The foreground region, perimeter pixels measured, edge orientation histograms, and texture aspects were all incorporated into the holistically imagined attributes. A total of 30 traits were collected, and the number of pedestrians traveling in all directions was estimated with Gaussian process regression. Some writers utilized approaches to decreasing dimensions. suggests that high-dimension holistic features be used, followed by dimensionality reduction utilizing kernel dimension reduction and the analysis of principal components[23].Instantly uses semi-supervised elastic network extrapolation (a sparse linear model) using a portion of the 129 holistic variables that were chosen. For crowd counting, many algorithms have been put out in the literature. Based on increasing appearance and motion data, earlier systems utilized a detection-style architecture that scanned a detector across two consecutive frames of a video sequence to estimate the number of pedestrians. have employed a similar detection-based methodology for counting pedestrians. In detection-based crowd-counting techniques, it is usual to presume that a crowd is made up of discrete entities that can each be identified by a particular detector. The drawback of these detection-based techniques is that the performance of the detector is strongly impacted by the occlusion of people in a crowded area or in a very thick crowd, which reduces the accuracy of the final estimation. People have suggested clustering the trajectories of monitored visual characteristics to count crowds in videos. For instance, it has used agglomerative clustering and a highly parallelized version of the KLT tracker to estimate the number of moving people. Basic picture characteristics have been monitored and probabilistically grouped into clusters that represent independently moving objects. These tracking-based techniques, however, are ineffective for calculating crowd sizes from specific still photos. Feature-based regression is conceivably the approach for crowd counting that is used the most. The key components of this kind of approach are foreground segmentation, foreground feature extraction (such as crowd mask edge count or texture characteristics), and crowd estimation using regression. Piecewise or linear functions are straightforward models that produce acceptable results. Ridge regression and the Gaussian process are other more sophisticated and efficient techniques. neural network and regression. Several data sources are used to estimate how many people are in a very crowded scene that can be seen in a single shot. In that paper, a dataset of fifty crowd photos with human annotations is presented. has kept up with the effort and calculated counts by combining data from several sources, namely interest locations. Low-confidence head detections, wavelet decomposition, GLCM features, and Fourier analysis has developed a support vector machine that produces counts for still pictures using the characteristics extracted from a pre-trained CNN. Recently, a method to count crowds in various settings was suggested based on CNN. They pre-trained a network first for particular scenarios. After receiving a test image from a fresh scene, to fine-tune the pre-trained network, they select similar training data based on the anticipated information and similarity in the density map. Their approach performs admirably on the majority of current datasets. However, their approach necessitates perspective maps for both the test scene and the training scene. Unfortunately, the availability of perspective maps restricts the practical use of such techniques in many crowd-counting applications. In conclusion, systemic methods rely on the assumption that worldwide image traits (holistic features) are most effectively employed for estimating a global measure (crowd size). Nevertheless, because of the broad spectrum of crowd dynamics, dispersion, and weight, crowd dimensions are difficult to measure. Techniques that are local and transitional try to address it.

2.2 Intermediate approaches

put forth an intermediary approach that employs blob-size histograms to fully indicate image features. The variety of both item sizes and how they appear in an area have been recorded utilizing the blob length stochastic and border alignment stochastic methods. Every blob's length is determined and utilized for categorizing it into a visual representation of a histogram bin, with each pixel weighted to the value it has on an image of density. The goal of the blob size histogram is to show the differences between the blobs that appear within an image; for instance, it is anticipated that chaos would fill the narrowest distribution container, while each person and tiny collections would slowly fill it. The model of regression learns the exact parameters of the association; nevertheless, by employing blob measurement distribution bins as imagined attributes, the algorithm is supposed to be able to distinguish between collections of people as well as individuals. Similar to the approaches outlined above, they work best in areas with little human habitation. Although overcameras were first proposed as a solution to the counting problem, it is unlikely that the most recent deployments of general-purpose cloud surveillance cameras will work with this method. The idea of head detection was first created by These methods work effectively in crowds where every person's face can always be seen on the camera, but they don't offer a complete solution to the group counting problem. The aforementioned techniques, which rely on detection-based computer systems, typically assume that there aren't many people nearby or that the particular camera is placed in the proper location. On the other hand, localization-based systems divide the image into various regions and try to regionally identify group members' parts. They also retrieved edge pixels with the positions of rotation using the edge detector [24]. The foreground area hides these pixels, enabling the background's limits to be disregarded. brought through a middle technique that fully identifies image characteristics by using blob-size histograms. The blob length stochastic and border alignment stochastic approaches have been used to capture the variation of both item sizes and how they appear in a region. Each pixel is weighted according to the value it contributes to an image of density, and the length of each blob is determined and used to classify it into a visual representation of a histogram bin. The blob size histogram's objective is to display the differences among the blobs that occur in an image; for example, it is expected that chaos would fill the smallest distribution container the fastest, while each individual and small collection would fill it gradually. The regression model determines the precise parameters of the algorithm and is intended to be able to discriminate between collections of people as well as individuals by using blob measurement distribution bins as imagined qualities. The edge angle histogram is created using eight bins with angles ranging from 0 to 180. The border orientation histogram "can distinguish pedestrian-generated edges, which are often vertical, from other scene features, such as noise, shadows, and automobiles. Several visual surveillance investigations lend credence to the claim. For instance, the edge orientation distribution and the blob's size histogram are used to represent an image, as described in the objective-oriented gradients (HOG), with the explicit objective of person recognition (even though their method is block-based and uses local standardization). To replicate the group size, it was calculated using both traditional regression and artificial neural network analysis. As a compromise between comprehensive and local features, this plan has been considered in the assessment. Eight bins between 0° and 180° are used to generate an edge angle histogram. The border orientation histogram "can discriminate edges generated by pedestrians, which are typically vertical, from other scene elements like noise, shadows, and automobiles. The assertion is supported by several studies on visual surveillance. For example, as described in the objective-oriented gradients (HOG) with the explicit objective of person recognition (despite the fact their method is block-based utilizing locals. The combination of Kong's features vector, which is employed for representing an image, is the edge orientation distribution and the blob's size histogram. To simulate the group size using each method, conventional regression and artificial neural network analysis were utilized. This strategy has been taken into consideration in the assessment as a compromise between comprehensive and local elements by using deeplearning[25].

2.3 Local approach

Local techniques for counting crowds utilize detectors and traits that are specific to specific individuals or collections of people in an image. The crowd estimation as a whole is the combination of each person's analyses for these categories. These strategies can be classified in general as follows: Detection-based approaches employ delineation computations using the head, face, or human sensors to figure out the approximate location of each individual within the environment. After that, there is a direct gathering process. Localization-based approaches split an image into various subregions before implementing regression-based determining approaches locally.Individual recognition of pedestrians works well in moderate crowds[26]. These approaches work most effectively in a sparse environment where the recognized component can be seen clearly. These strategies aren't addressed completely here since the topic of this study involves congested and obstructed configurations. It offers an analysis of the various pedestrian detection methods presently in common use[27].Crowd segmentation is a substitute for recognizing pedestrians. In this approach, the rough position of people in the environment is calculated in a bid to explain the observed visual attributes. stipulated an approach for human separation within a model-based probabilistic architecture and suggested that this knowledge could have been obtained from the foreground mask. Each of the four ellipses that comprised the individual's 3D model had adjustable parameters[28].Technique are employed to repeatedly and non-exhaustively search the range of solutions within elevated likelihood areas to figure out the most suitable alternative. Because of the substantial complexity of the outcome space, the method's ability to perform actual time segments in situations with more than 10–15 residents is a constraint. Furthermore, as the significance of various position models grows, it might be possible that this data will be obscured from sight. Constructing an ensemble of different Bernoulli forms for capturing foreground people using a training set of images, for instance, indicated an example-based technique that expanded this to a multi-camera system[29]. employed an example-based approach whereby shape descriptors were used to communicate blobs of a small size instead of explicit shape representations. Fourier descriptors are used to encode the approximate shape of a blob, with the highest-frequency coefficients being disregarded since they make up just a fraction of the total blob shape. To assess fresh data by interpolation using K closest neighbors (KNN) regression, the initial dataset comprised groups of commuters organized in various arrangements. In groups with sizes 1–6, the method was placed on exhibit. The quantity of training data that is provided limits this approach. The example-based technique is inadequate as the group's membership rises due to the challenge of gathering sufficient training information for every potential pedestrian combination. Alternative crowd segmentation techniques, such as maximization of expectations and 2D models, have been developed, but they are not intended to function in crowds regardless of size. Other approaches have attempted to take advantage of tracking or symmetry[30]. Similar to the aforementioned methods, they perform best in sparsely settled regions. While over-cameras were originally presented as a means of improving the counting problem, the latest deployments of general-purpose cloud surveillance cameras are not likely to prove compatible using this technique. The concept of head detection was initially developed by These strategies perform well in crowds where each individual's face is always visible on the camera, but they lack a comprehensive answer to the group counting challenge. The previously mentioned methods, which rely on detection-based computer programs, usually start with the assumption that there are not many individuals near or that the specific camera is situated in an appropriate spot. Localization-based solutions, on the other hand, partition the image into different regions and attempt to identify units within the group regionally. To count crowds in movies, several people have suggested clustering the trajectories of recorded visual attributes. For instance, has employed a highly parallelized version of the tracker with agglomerative clustering to estimate the number of moving people. To depict independently moving objects, essential visual features have been tracked and probabilistically aggregated into clusters. However, crowd sizes cannot be determined from specific still photographs using these tracking-based algorithms. Possibly the most popular method for crowd counting is feature-based regression. Foreground segmentation, foreground feature extraction (such as crowd mask edge count or texture properties), and crowd estimate using regression are the primary elements of this type of approach. Simple models that deliver respectable results include piecewise or linear functions. The Gaussian and ridge regression procedure is another more advanced/effective method. Regression and neural networks. The number of individuals in an extremely crowded scene that can be visible in a single photo is estimated using a variety of data sources. A dataset of 50 crowd photographs with human comments is offered in that publication. has continued the project and combined information from several sources, especially interest locations, to calculate counts. Wavelet decomposition, features, low confidence head detections, and Fourier analysis. has created a support vector machine that uses attributes taken from trained output counts for still images. A technique for counting crowds in diverse contexts has recently been proposed. They first pre-trained a network for specific scenarios. following the test picture of a recent scene, Based on the anticipated information and similarity in the density map, they choose similar training data to fine-tune the pre-trained network. On the vast majority of recent datasets, their method performs superbly. However, their method calls for perspective maps for both the training and test scenes. In many crowd-counting applications, however, the actual use of such approaches is constrained by the accessibility of perspective maps. The idea behind systemic approaches is that global image attributes (holistic aspects) may be used most effectively to estimate a global measure (crowd size). Crowd dimensions are challenging to quantify because of the wide range of crowd dynamics, dispersion, and weight. It is attempted to be addressed by local and transitional techniques.

A shifting keypoint cluster was proposed to carry out group localization[31]. With this method, important points inside an image have been identified using SURF[32]. These stationary areas are subsequently overlooked as a result of the light flow concealing these areas. An arrangement of the relevant moving components.

Chapter 3

Methodology

First, we apply the technique outlined into infer a viewpoint normalization map[33]. We extract low-level imaging features, such as local foreground, edges, and texture features, from each cell region using a set of training photos. Before all local intermediate feature vectors are concatenated into a single ordered (location-aware) feature vector, local features from each cell are used to create a local intermediate feature vector in step three. Using the single concatenated feature vector and the vector, each element of which is the actual count in each region, as a training pair, a multi-output regression model based on multivariate ridge regression is built. Features are taken from a fresh test frame and transferred to the learned regression model to produce a structured output that calculates the crowd size.

One concatenation vectors of attributes and the vector, every element of which represents the real number of features for every area, are employed to build a multi-output model of regression using a multivariate approach using ridges. by providing a test pair. Variables from an entirely new testing frame have been added to the trained regression framework to provide structured results which calculate the size of the population. To offer a structured input that estimates the crowd size, attributes from an entirely novel test frames are added to a learnt predictive models in the below figure (see Figure 3.1)by offering a test couple. The trained regression framework now includes variables from a brand-new testing framework to produce structured findings that determine population size. A learned model for prediction gets improved with features from entirely novel test.

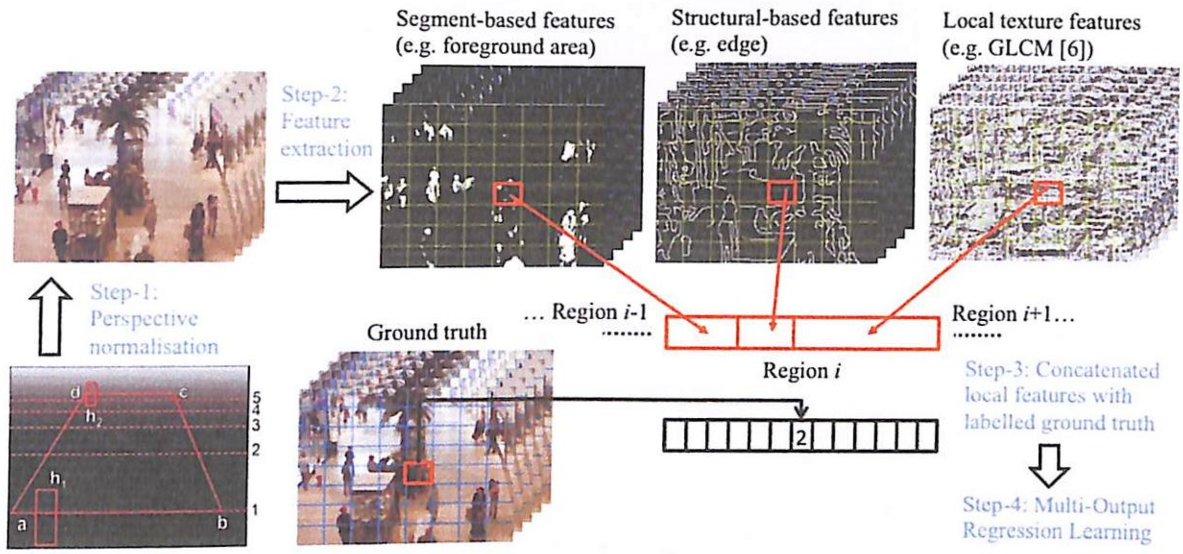

Figure 3.1: A multi output regression technique for localized crowd counting using feature mining

3.1 Representation of the Features

We initially divide the frame into K cell sections using the training video frame I, where i = 1, 2,... N, and N represents the total number of training frames (see Step-3 in Figure 1). Then, from each cell region j, we extract low-level imaging the above equation (see equattion ??)and (see equation ??) combine them to create an intermediate feature vector xi Rd. Additionally, we create a multi-dimensional output vector by concatenating the localized labeled ground truth u j from each cell region.

$$\mathbf{x}_i = \left[\mathbf{z}_i^1, \mathbf{z}_i^2, \ldots, \mathbf{z}_i^{K-1}, \mathbf{z}_i^K\right], \quad \mathbf{y}_i = \left[u_i^1, u_i^2, \ldots, u_i^{K-1}, u_i^K\right] \tag{3.1.1}$$

$$\{(\mathbf{x}, \mathbf{y})\}_i, i = 1, 2 \ldots N \tag{3.1.2}$$

- Foreground area, total perimeter pixel measure, perimeter-area proportion, and the histogram of the perimeter edge introductions are segment-based characteristics.

- Structural features consist of all the edge pixel measures, the histogram for edge introductions, and Minkowski, a dimension.

- Local texture features include the gray level co-occurrence matrix (GLCM).

It should also be mentioned that before the extraction of features, every image is converted to grayscale. In addition, features have been transformed into [0, 1] and perspective normalized using the method described in [2].

3.2 Model for Multiple Output Regression

We utilize the regression ridge function to acquire a multi-output model for regression [34]. A ridge regression function trains one input mapping of its conventional configuration. In the present case, we tweak it to address a multi-output logistic training issue that involves concurrent targeted crowd counts across several geographic unit places. The rationale for employing ridge regression is that, in comparison with ordinary least-squares error minimizing, the framework offers greater durability when dealing with the multiple convergence problem. In standard methods of regression, like linear regression, the least-squares method is used. For the recognition of faces in other programs, regression with ridges is used. The first attempt at utilizing it for crowd analysis is this one. Additionally, given (X_i, Y_i).

The following is a representation of the presentation of multivariate ridge regression using observations and target vectors:

$$\min \frac{1}{2}||\mathbf{W}||_F^2 + C\sum_{i=1}^{N} ||\mathbf{y}_i^T - \mathbf{x}_i^T\mathbf{W} - \mathbf{b}||_F^2 \tag{3.2.1}$$

$$\mathbf{W} \in \mathbb{R}^{d\times K} \text{ and } \mathbf{b} \in \mathbb{R}^{1\times K} \tag{3.2.2}$$

indicating the Frobenius norm. The weighting of the matrix W plays a major role when assessing the value of local features or facilitating feature exchange. We explicitly build a model for every limited cell to simultaneously weigh the characteristics assembled from these localized cells alongside the other cell areas in the graphic. As indicated in Equation (1) above, the spliced feature vector xi in every jth cellular section in the images weights the jth row in vector W to the total number of cells estimated in the associated localized cell area, or the jth entry of yi. About the resulting error of all cell regions being punished together using the Frobenius-norm and the feature vector xi that consists of characteristics from the second jth cell geographical area along with other cell areas in the image, like a model of regression might profit from neighborhood significance extraction along with, especially, from characteristic exchange inside the image's overall room for the confined take an estimation of the density of a specific area. In this section, we go into greater detail regarding error minimization. The above equation (1) has been modified exactly as follows:

$$\min M(\theta) = \operatorname{tr}\left(\frac{1}{2}\theta^T Q\theta + P^T\theta\right) \tag{3.2.3}$$

The Structural Support Vec- for Machines, which was used for stance estimates [35]and finding objects , serves as an option to the multi-output recessive model. Because of the simplicity of utilization, multidimensional regression with ridges is frequently utilized.

Chapter 4

Dataset and Results

4.1 Experiments

contains information about the two datasets, and displays the adequate frames. The newly released Mall dataset and ucsd dataset was collected utilizing a publicly accessible security camera in a shopping mall(see Figure 4.1) and with challenging lighting situations and glass surfaces, compared with the ucsd datasetwhere the images were gathered from a campus setting using a hand-held camera. The Mall dataset additionally encompasses an increased number of crowd concentrations, from limited to overcrowded, in addition to different activity patterns incremaining and moving crowds) in a broader spectrum of lighting conditions at different times of the day.The Mall dataset also goes through greater amounts of perspective distortion than the ucsd dataset, which results in bigger variation between it and the visibility of items at different tiers of the scene, along with frequent shadowing issues caused by scene components.By supplying a test pair, one concatenation vectors of attributes and the vector, each of which elements reflects the actual number of features for each area, are used to develop a multi-output regression model using a multivariate technique employing ridges. The trained regression framework now includes variables from a brand-new testing framework to produce structured findings that determine population size. To provide a structured input that gauges the size of the crowd.

contains information about the two datasets, and displays the adequate frames. The newly released Mall dataset and ucsd dataset was collected utilizing a publicly accessible security camera in a shopping mall(see Figure 4.1) and with challenging lighting situations and glass surfaces, compared with the ucsd datasetwhere the images were gathered from a campus setting using a hand-held camera. The Mall dataset additionally encompasses an increa=p.[]sed number of crowd concentrations, from limited to overcrowded, in addition to different activity patterns incremaining and moving crowds) in a broader spectrum of lighting conditions at different times of the day.The Mall dataset also goes through greater amounts of perspective distortion than the ucsd dataset, which results in bigger variation between it and the visibility of items at different tiers of the scene, along with frequent shadowing issues caused by scene components.By supplying a test pair, one concatenation vectors of attributes and the vector, each of which elements reflects the actual number of features for each area, are used to develop a multi-output regression model using a multivariate technique employing ridges. The trained regression framework now includes variables from a brand-new testing framework to produce structured findings that determine population size. To provide a structured input that gauges the size of the crowd.

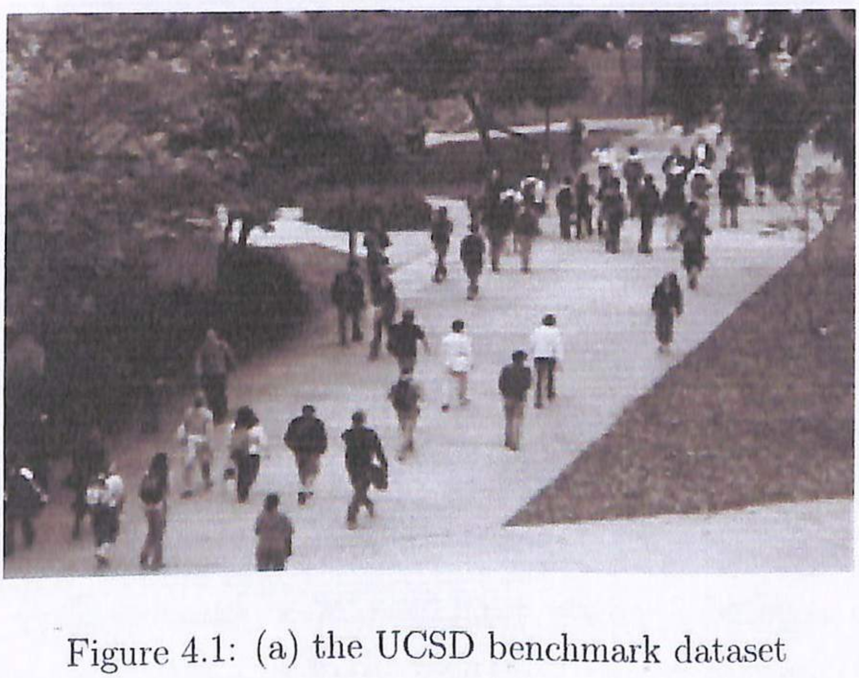

Figure 4.1: (a) the UCSD benchmark dataset

4.2 Comparative Analysis of Single Global Regression Techniques

The newly released Mall dataset was collected utilizing a publicly accessible security camera in a shopping mall with challenging lighting situations and glass surfaces, compared with the images where the images were gathered from a campus setting using a hand-held camera. The Mall dataset additionally encompasses an increased number of crowd concentrations, from limited to overcrowded, in addition to different activity patterns that remain and move crowds in a broader spectrum of lighting conditions at different times of the day. it also goes through greater amounts of perspective distortion than the U, which results in bigger variation between it and the visibility of items at different tiers of the scene, along with frequent shadowing issues caused by scene components such as stalls and plants in the interior that are in the way of approach.

4.3 The assessment of Our Model's Local Features Extraction

Our multi-output system for local crowd counting has the advantage of leveraging the importance of specific features to support crowd density estimations.various types of features can be of greater importance to various extents in an area when evaluating a specific local crowd density. An image of a recent event, To fine-tune the pre-trained network, they select similar training data based on the anticipated information and similarity in the density map. They have excellent results using their strategy on the vast majority of current datasets. However, for both the training and test situations, their approach requires perspective maps. However, the availability of perspective maps in many crowd-counting applications limits the actual usage of such methods. According to systemic techniques, it is possible to estimate a global measure (crowd size) most effectively by using global picture qualities (holistic features). Due to the wide range of crowd dynamics, dispersion, and weight, crowd dimensions are difficult to calculate. Local and interim techniques are used to try to remedy it. Particularly plots demonstrate that the edge orientations 60 (which mainly lead to shoulder sides) within the away-from-camera cell 43 showed increased significance. comparable with different edge orientations; nevertheless, The advantage of our multi-output method for local crowd counting is that it may make use of the significance of particular variables to support crowd density estimates. When assessing a particular local crowd density, different feature types may be of higher significance to different degrees in a given area, as shown in a picture of a recent occurrence. Based on the anticipated information and similarity in the density map, they chose similar training data to further refine the pre-trained network. On the vast majority of the available datasets, they achieve good outcomes using their approach. However, their strategy needs perspective maps for both the training and test scenarios. However, the actual implementation of such techniques is constrained by the prevalence of perspective maps in crowd-counting applications. Systemic methods suggest that the edge's orientation of 90° in Close-to-Camera Cell 6 received an additional weight. This edge orientation mainly correlates to the torso surfaces. At the bottom of Figure 4, we offer two representative patches of the image and the associated edges that were recovered from cells 6 and 7. Our multi-output technique for local crowd counting has the benefit of taking advantage of the significance of particular traits to support crowd density estimates. When assessing a particular local crowd density, as its different feature types may be of higher significance to different degrees in a given location. a photo of a recent occasion, They choose comparable training data based on the predicted information and similarity in the density map to fine-tune the pre-trained network. On the majority of recent datasets, they have achieved great outcomes using their approach. However, their method necessitates perspective maps for both the training and test environments. The use of such techniques is, however, constrained by the prevalence of perspective maps in crowd-counting applications. Systemic methods indicate that they give an in-depth look at the various edge orientations in various cells. The results indicate that the values found by our model match what we suspected, i.e., that distinct features, such as shoulders and torso borders, are regarded as differential for determination across the deepest parts of a scene.

4.4 Evaluation of Regional Sharing of Information

We utilized the Mall dataset as an investigation instance and chose three cells as the ones to investigate whether adjacent cells contribute to the count estimation to demonstrate that spatially restricted parts in an image are likely to communicate the data. By adding the total weights of adjacent cells that contribute to the overall count of the cell that interests us, we can determine the level of information sharing or evidential support in the image area.The amount of contribution or information exchange can be transformed into a visualization called a heat map. Additional data disappeared among adjacent cells the closer they got to the designated cell area. This result shows that to get a more accurate counting estimation, our algorithm has the capability of requesting evidence from more cell areas.

4.5 Comparative Analysis for the Multiple Localized Regression Model

Our multi-output regression technique performed superiorly to a single local regression model (MLR).Although the MLR was created to address the drawbacks of global regression techniques, it is significant to note that its accuracy is even lower compared to that of the two global regression models. This outcome underlines the significance of utilizing the relationship among features across areas and exchanging information among areas to obtain more precise crowd qualifiers since the MLR framework evaluates the importance of different attributes in various localized areas in a way that's comparable to our strategy. The local measurements are too chaotic and volatile for them to be reliant upon solely to forecast densities absent this data sharing, as is rendered possible using the variety of structural input regression techniques established in the present investigation. Relative to the MLR model, our single model-based regression method is more operationally scalable; for instance, MORR is 3-5 times more efficient for both testing and training than MLR.offers additional details regarding MLR and MORR evaluations and instruction programs and as you can see (see Figure 4.2)

Method	Features Level		Learning Level		UCSD [6]		
	Global	Local	Global	Local	mae	mse	mde
RR [22]	✔	–	✔	–	2.25	**7.82**	0.1101
GPR [6]	✔	–	✔	–	**2.24**	7.97	0.1126
MLR [23]	–	✔	–	✔	2.60	10.1	0.1249
MORR	–	✔	✔	–	2.29	8.08	**0.1088**

Figure 4.2: Performance comparison between different methods and our multi-output ridge regression (MORR) model

4.6 Localized Counting Reliability Assessment

To assess the efficacy of MORR and MLR, we selected two busy locations (directly in front of two businesses) from the Mall dataset to demonstrate the value of the suggested MORR in localized counting. These chosen regions are shown in a figure with the results presented from localized crowd counting. Our MORR provided a more precise localized count than MLR, in addition to taking less time for preparation and evaluation. The results also point to the importance of local data exchanges. Using several different data sources, one snapshot is estimated. In that article, a dataset of 50 crowd photos with user comments is provided. carries on the project and combines data from several sources, particularly interest sites, to determine counts. Low-confidence head detections, wavelet decomposition, features, and Fourier analysis take attributes from a trained output count for still images to generate a support vector machine. A method for counting crowds in various circumstances has recently been put forth. A network was initially trained in advance for particular scenarios. After the test image of a recent event, to fine-tune the pre-trained network, they select similar training data based on the anticipated information and similarity in the density map. The majority of recent datasets show that their technique operates flawlessly. However, for both the training and test situations, their approach requires perspective maps. However, the availability of perspective maps in many crowd-counting applications limits the actual usage of such methods. According to systemic techniques, it is possible to estimate a global measure (crowd size) most effectively by using global picture qualities (holistic features). Due to the wide range of crowd dynamics, dispersion, and weight, crowd dimensions are difficult to calculate. Local and interim techniques are used to try to remedy it.

For group localization, a shifting keypoint cluster was suggested. Important details inside an image that have been discovered by the stationary areas using this technique are then missed due to the light flow. Agglomerative clustering is used in a parallelized version of the tracker to calculate the population in motion. Essential visual properties have been tracked and probabilistically aggregated into clusters to represent independently moving objects. However, these tracking-based algorithms cannot estimate crowd sizes from specific still images. The technique operates flawlessly. However, for both the training and test situations, their approach requires perspective maps. However, the availability of perspective maps in many crowd-counting applications limits the actual usage of such methods. covering up these spaces.configuration of the necessary moving parts The technique of feature-based regression for crowd counting may be the most common. The main components of this kind of strategy are foreground segmentation, foreground feature extraction (such as crowd mask edge count or texture attributes), and crowd estimation utilizing regression. Piecewise or linear functions are simple models with decent outcomes. Another more sophisticated and effective technique is the Gaussian and ridge regression approaches. network simulation and regression. The maximum number of people in a densely populated area that may be seen.

Chapter 5

Conclusions and future work

5.1 Conclusions

We proposed a single, multi-output model of regression that can determine crowds in particular places. Our approach utilizes a single joint regressor that takes numerous spliced specialized imaging characteristics as input for learning spatially specialized population counts, typically with numerous outputs, rather than producing several specialized regressors as used by current approaches. Due to its inherent ability for feature extraction according to shifting crowd circumstances given in different local geographic cells, our approach beats various localized regressors on an extremely difficult shopping mall dataset. Comparing experimental results shows the value of our strategy. Future research will focus on improving the density of crowds to determine accuracy by taking into account the population structure's dynamic and temporal separation. This will make it easier to let go of any predetermined localized cell region definition and enable additional details. efficient population-responsive count models that constantly represent changing crowd structures for choosing features and representations. Yet, it additionally performs well if contrasted with similar single-worldwide regressor-based population counting models. A single-feed active as well as semi-supervised regression is used to enable crowd estimation with only a few marked sample images by using the fundamental distribution pattern of crowd patterns, which provides easily accessible enormous amounts of unlabelled information compared to most existing crowd calculation examines, which rely on comprehensive annotations for the training of models. Furthermore, we proposed an entirely novel system and concept for transfer counting. We explored how details from earlier scenes can be used to minimize the quantity of work required to start crowd counting in a new scene if marked information is absent. It's very useful in real life. We used the assumption that the original and target information are equivalent in the transfer counting technique, having an analogous depiction of a manifold. Future research will look at ways to reduce this presumption using automated source-target relevancy assessments. Counting individuals can help with effective crowd control. By looking at crowd demographics and flow patterns, event planners or security professionals can employ strategies that increase crowd flexibility, minimize bottlenecks, and maintain a fluid flow among people inside the venue. Allocating resources is made easier by accurate crowd counting. The event planners can estimate the number of attendees and then distribute supplies like furniture, beverages, and amenities by that amount. This guarantees that the infrastructure and amenities can satisfy the needs of the people in attendance. Individual counts at busy locations and establishments, such as arenas, shopping malls, or transportation hubs, can provide information on "real-occupant tracking." Managers of facilities can control admission to specific domains and track entry using this information. and exit points, keep the area hospitable and safe for all visitors to use. Planning for disasters: People are the most crucial component of emergency preparedness. When there are dense crowds, emergency personnel can organize and execute evacuations by having a complete list of everyone present. This guarantees everyone's safety regardless of a fire, an earthquake, or a danger from terrorism. Crowd analytics: Information gathered through person counts may be used to evaluate the trends and behaviors of the crowd. Event organizers and advertisers can make informed decisions about budget allocations, tactics for involving spectators, and targeted advertising by studying crowd dynamics. With congested metropolitan locations like transit hubs, crowd tracking can aid with traffic management. Authorities may change traffic signs, point vehicles, and provide additional transit service by keeping track of how many people enter or leave particular areas to reduce congestion. using market and advertising intelligence The statistical analysis of businesses and advertising can both benefit from data on the number of people. Shops can use footfall traffic data to understand consumer behavior, lay out their stores, and evaluate the effectiveness of their marketing campaigns. This knowledge may be used to make data-driven business decisions while enhancing client experiences. Healthcare with Social Distinction: For those that are concerned about everyone's health when there are disease outbreaks or epidemics, counting could help facilitate the implementation of measures for societal isolation. Governments can reduce the risk of virus transmission and enforce adherence to safety requirements by keeping an eye on the number of people and putting in place occupational restrictions. Using prediction modeling in management Long-term population-qualifying statistics can be utilized for planning and forecasting. By studying historical limit behaviors, urban developers, event planners, or transportation officials can forecast future crowd sizes, plan the construction of infrastructure, improve transit services, or advance the general growth of cities.

5.2 Future work

We will eventually expand on our existing approach to handle surveillance scenarios that involve multiple cameras. To improve the accuracy of head detection, we may employ cutting-edge classification approaches and posture classification methods in particular. In addition, we will incorporate monitoring methods into our approach to managing people-flow matters. For instance, methods of tracking can be employed for collecting human movement information when we've determined our head's location. Therefore, we can count the number of people moving in various directions. On the other hand, crowded areas include theaters, sports arenas, events, political speeches, and meetings. just one photo. A "public count" method is used to estimate the number of free items (such as people, cars, cells, etc.) in scenarios where the scene is unrestricted in and of itself. Crowd counting is currently generating a lot of interest due to its critical applications in areas including consumer behavior, traffic control, public safety, and cell counts. Numerous methods have been used to comprehend the significance of carry forecasting, particularly the use of machine learning, which is more effective for several applications such as testing hypotheses, image categorization, and various kinds of co-time series. Crowd counting is used to determine how many people can be seen in photographs and videos. It is a significant subject with many real-world applications, including video surveillance (handling complaints in a specific location or identifying densely packed audiences), security management (counting the number of people entering or leaving certain areas), web multimedia (estimating the crowd size for tweet image launching within a crowded scenic spot), and more. It's a crucial topic in many real-world applications, including surveillance footage (such as managing complaints for a specific area or identifying audience members in a crowded space), security monitoring (such as counting the number of people who enter and leave only a few zones), and the internet (such as crowd size forecasting over tweet shootings). against a densely populated scenic backdrop), a demographic census is a crucial task that could effectively make use of population facts. There are still many practical applications for load prediction, including forensic search, catastrophe management, the building of large areas, information gathering and analysis, and the use of virtual worlds. Images of people's behavior can be studied using monitoring techniques like crowd counting. To prevent calamities, crowd classifiers can identify crowding. Building designs for public spaces can be improved using the findings of studies on crowd size or mobility. Crowd counting could be used to collect data and analyze it to gauge popular interest. The identification of faces using crowd-counting techniques may advance present technologies. As part of our effort, we set out to build a system for counting people in extremely thick crowds. This can help with a variety of global issues. It might assist management with safety, for example. Unusual actions Detection The importance of demonstrations or rallies politically It was designed and studied to count people in extremely crowded areas using an image-based approach. We gave ourselves the assignment of developing an image-based system that can recognize densely populated places and may have an effect in numerous locations. The main objective of crowd counting is to accurately track people in crowd photographs. Modern methods use multi-column CNN designs to regress density maps of crowd photographs to handle the scale or perspective modifications that frequently occur in crowd photos. The reception areas of various columns reacted differently to varying sizes of people. The primary objective is to develop a system that can precisely count large, dense crowds of people using population pictures. Following are some categories for crowd counting's goals and objectives: Security first: Ensuring that everyone is safe is one of the main goals for conducting a human census in places with large populations. By carefully counting the number of people present in a crowded place, officials can monitor the degree of crowding and take action to prevent crowding, stampedes, or other potentially dangerous circumstances. Administration of crowds: Counting people makes crowd control more effective. Event planners or security professionals may use strategies that increase crowd adaptability, eliminate bottlenecks, and guarantee seamless movement among people inside the crowd. by analyzing the demographics and movement patterns of the audience. Allocating resources is made easier by accurate crowd counting. The event planners can estimate the number of attendees and then distribute supplies like furniture, beverages, and amenities based on that number. This guarantees that the infrastructure and amenities can satisfy the needs of the people in attendance.

Bibliography

[1] Ya Chang, Changbo Hu, Rogerio Feris, and Matthew Turk. Manifold based analysis of facial expression. Image and Vision Computing, 24(6):605–614, 2006.

[2] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In 2008 19th international conference on pattern recognition, pages 1–4. IEEE, 2008.

[3] Tao Zhao, Ram Nevatia, and Bo Wu. Segmentation and tracking of multiple humans in crowded environments. IEEE transactions on pattern analysis and machine intelligence, 30(7):1198–1211, 2008.

[4] 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.

[5] David Ryan, Simon Denman, Clinton Fookes, and Sridha Sridharan. Crowd counting using multiple local features. In 2009 digital image computing: techniques and applications, pages 81–88. IEEE, 2009.

[6] Antoni B Chan, Zhang-Sheng John Liang, and Nuno Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In 2008 IEEE conference on computer vision and pattern recognition, pages 1–7. IEEE, 2008.

[7] Siu-Yeung Cho, Tommy WS Chow, and Chi-Tat Leung. A neural-based

crowd estimation by hybrid global learning algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(4):535–541, 1999.

[8] Weina Ge and Robert T Collins. Marked point processes for crowd counting. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 2913–2920. IEEE, 2009.

[9] Yoel Haitovsky. On multivariate ridge regression. Biometrika, 74(3):563–570, 1987.

[10] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6):610–621, 1973.

[11] D Helbing and A Johansson. Eidgenössische technische hochschule. Pedestrian, crowd and evacuation dynamics. Encyclopedia of Complexity and Systems Science, 2009.

[12] Dan Kong, Douglas Gray, and Hai Tao. A viewpoint invariant approach for crowd counting. In 18th International Conference on Pattern Recognition (ICPR'06), volume 3, pages 1187–1190. IEEE, 2006.

[13] Dan Kong, Douglas Gray, and Hai Tao. Counting pedestrians in crowds using viewpoint invariant training. In BMVC, volume 1, page 2. Citeseer, 2005.

[14] Anthony C Davies, Jia Hong Yin, and Sergio A Velastin. Crowd monitoring using image processing. Electronics & Communication Engineering Journal, 7(1):37–47, 1995.

[15] Antoni B Chan and Nuno Vasconcelos. Bayesian poisson regression for crowd counting. In 2009 IEEE 12th international conference on computer vision, pages 545–551. IEEE, 2009.

[16] Nan Dong, Fuqiang Liu, and Zhipeng Li. Crowd density estimation using sparse texture features. J. Convergence Inf. Technol., 5(6):125–137, 2010.

[17] Wenhua Ma, Lei Huang, and Changping Liu. Crowd estimation using multi-scale local texture analysis and confidence-based soft classification. In 2008

second international symposium on intelligent information technology application, volume 1, pages 142–146. IEEE, 2008.

[18] AN Marana, L da F Costa, RA Lotufo, and SA Velastin. On the efficacy of texture analysis for crowd monitoring. In Proceedings SIBGRAPI'98. International Symposium on Computer Graphics, Image Processing, and Vision (Cat. No. 98EX237), pages 354–361. IEEE, 1998.

[19] Robert M Haralick. Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5):786–804, 1979.

[20] Aparecido Nilceu Marana, L Da Fontoura Costa, RA Lotufo, and Sergio A Velastin. Estimating crowd density with minkowski fractal dimension. In 1999 IEEE international conference on acoustics, speech, and signal processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), volume 6, pages 3521–3524. IEEE, 1999.

[21] Li Xiaohua, Shen Lansun, and Li Huanqin. Estimation of crowd density based on wavelet and support vector machine. Transactions of the Institute of Measurement and Control, 28(3):299–308, 2006.

[22] Hidayah Rahmalan, Mark S Nixon, and John N Carter. On crowd density estimation for surveillance. 2006.

[23] Vu Pham, Phong Vo, Vu Thanh Hung, et al. Gpu implementation of extended gaussian mixture model for background subtraction. In 2010 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pages 1–4. IEEE, 2010.

[24] Duan-Yu Chen and Kuan-Yi Lin. A novel viewer counter for digital billboards. In 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 653–656, 2009. doi: 10.1109/IIH-MSP.2009.211.

[25] Mohamed Ibrahim and Cemil Turan. Deep learning in crowd analysis: A survey of current research trends. In 2023 17th International Conference on Electronics Computer and Computation (ICECCO), pages 1–3. IEEE, 2023.

[26] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9): 1627–1645, 2010. doi: 10.1109/TPAMI.2009.167.

[27] Lan Dong, Vasu Parameswaran, Visvanathan Ramesh, and Imad Zoghlami. Fast crowd segmentation using shape indexing. In 2007 IEEE 11th International Conference on Computer Vision, pages 1–8, 2007. doi: 10.1109/ICCV.2007.4409075.

[28] Simon Denman, Clinton Fookes, and Sridha Sridharan. Improved simultaneous computation of motion detection and optical flow for object tracking. In 2009 Digital Image Computing: Techniques and Applications, pages 175–182, 2009. doi: 10.1109/DICTA.2009.35.

[29] A.N. Marana, L. Da Fontoura Costa, R.A. Lotufo, and S.A. Velastin. Estimating crowd density with minkowski fractal dimension. In 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), volume 6, pages 3521–3524 vol.6, 1999. doi: 10.1109/ICASSP.1999.757602.

[30] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. Understanding collective behaviors: Learning a mixture model of dynamic pedestrian-agents. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2871–2878, 2012. doi: 10.1109/CVPR.2012.6248013.

[31] Keju Zu, Fuqiang Liu, and Zhipeng Li. Counting pedestrian in crowded subway scene. In 2009 2nd International Congress on Image and Signal Processing, pages 1–4, 2009. doi: 10.1109/CISP.2009.5303594.

[32] Vu Pham, Phong Vo, Vu Thanh Hung, and Le Hoai Bac. Gpu implementation of extended gaussian mixture model for background subtraction. In 2010 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pages 1–4, 2010. doi: 10.1109/RIVF.2010.5634007.

[33] Matthew B Blaschko and Christoph H Lampert. Learning to localize objects with structured output regression. ECCV (1), 5302:2–15, 2008.

[34] Ke Chen, Chen Change Loy, Shaogang Gong, and Tony Xiang. Feature mining for localised crowd counting. In Bmvc, volume 1, page 3, 2012.

[35] Wenhua Ma, Lei Huang, and Changping Liu. Crowd density analysis using co-occurrence texture features. In 5th International Conference on Computer Sciences and Convergence Information Technology, pages 170–175. IEEE, 2010.