INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 12 | DECEMBER - 2023 SJIF RATING: 8.176 ISSN: 2582-3930

Anaomaly Detection at Crowded Places

Prachi Pancholi
Department of Computer Science and Engineering
Chandigarh University
Mohali, India
prachipancholi06@gmail.com

Priyanshu Singh
Department of Computer Science and Engineering
Chandigarh University
Mohali, India
ks.priyansu@gmail.com

Shaifali Sharma
Department of Computer Science and Engineering
Chandigarh University
Mohali, India
shafalii.e13752@cumail.in

Abstract

As research has indicated, the population is tremendously increasing and the numbers are expected to reach approximately 9.7 billion by the year 2050, a significant increase from the 7.4 billion recorded in 2016. This continuous population growth has underscored the critical importance of crowd detection, making it a pivotal tool in a wide array of studies and applications, including crowd counting, urban planning, crime detection, and visual surveillance. This combination of capabilities opens the door to numerous novel studies and applications. For instance, crowd density detection becomes invaluable during times of pandemics, such as the COVID-19 crisis. It allows for the proactive monitoring and management of overcrowding situations, ensuring compliance with safety regulations, and maintaining proper social distancing measures.

The primary objective of crowd density detection is to identify and quantify the congregation of individuals within video footage, subsequently recognizing discernible patterns and generating output results based on these observed patterns.

Enhancing accuracy in crowd detection involves incorporating various complementary techniques alongside Convolutional Neural Networks (CNN). These methods include applying Haar filters to video frames and integrating Multi-Scale Convolutional Neural Networks (MSCNNs). The choice of approach for crowd detection and classification depends on the crowd's density being observed. In densely populated areas, the focus often shifts away from individual face detection towards analyzing crowd behavior and tracking crowd movements.

This review paper provides a comprehensive analysis of crowd detection and classification, encompassing a range of general techniques for crowd analysis and monitoring.

Keywords— Computer Vision (CV), Crowd Density estimation, detection, CNN, Crowd Detection, invisible reputation.

1. INTRODUCTION

In recent years, the growing population has piqued the interest of researchers in the fields of reliability, surety and CV for analyzing crowd mobility and behaviors to reduce load on human. The increase in street crowds has heightened the risk of accidents, mass panic, and crowd rushes, making public safety a more pressing concern. Figure 1 depicts some instances of crowd disasters during large-scale events

Figure1: (a) Mina, a city located in Saudi Arabia. (b) The Love Parade disaster in Duisburg. (c) The water festival at Phnom. (d) Bombing at Boston in Massachusetts, United States.[1]

There are several methods for detecting anomaly in crowd which includes detection-based method. Detection-based methods primarily focus on extracting low-level features, such as identifying faces, facial features, or complete human bodies. However, in crowded public places, it's often impractical to rely on facial recognition, especially considering the limited resolution of surveillance cameras. Consequently, detection-based algorithms may not be as effective in these scenarios.

Regression-based models offer a solution to the limitations of detection-based approaches. They begin by selecting significant patches from the image, cropping them, and subsequently extracting low-level features. This approach allows for a more versatile analysis.

Another approach is density estimation-based methods. These methods start by creating density maps for objects within the scene. Then, density maps are created. Alternatively, random

forest regression can be employed to capture non-linear mappings.
CNN-based methods rely on the robustness of convolutional neural networks (CNNs). Instead of focusing on image patches, these methods construct an end-to-end regression framework using CNNs. This directly gives the count estimation when an input is given. CNNs are renowned for their effectiveness in regression and classification tasks, and they have demonstrated their capability in generating density maps

1.1 Significance of model

In highly crowded places, recognizing individuals' faces can be challenging. Therefore, there is a growing need for surveillance systems to detect, track, and analyze crowd behavior. These analyses are crucial for safety purposes. Mass gatherings, including rallies, protests, and social events, often lack the manpower required for real-time monitoring. Continuous human monitoring is not always effective, as it can be prone to errors and is both tiring and cumbersome. Consequently, there is a significant importance in developing a model capable of detecting crowd anomalies to prevent potential mishaps.

1.2 Objective of review

The end goal of this survey paper is to identify and highlight cutting edge technologies present to detect and identify crowd anomaly. It emphasizes on the existing technologies, there disadvantages and the future scope of new evolving algorithms. It also points out the challenges and problems faced while analyzing video frame by frame

2. EXISTING APPROACHES

[I] Human detection- The process involves subtracting background frames to track human motion, followed by the application of optical flow and spatio-temporal filtering. To classify objects, shape and motion features are extracted from video frames. However, these methods can be influenced by external factors, affecting the results. To address these challenges, methods for considering optical flows include Horn and Schunck's approach, while for accurate results under varying illuminations, the pyramidical implementation of Lucas-Kanade's method is utilized. Additionally, sliding window detectors are employed along with Support Vector Machines (SVM) to enhance accuracy.

A. Spatio-temporal filtering- Digital images and videos undergo compression to eliminate spatial, temporal, and visual redundancies. While this compression results in more efficient storage and transmission, it disrupts the pixel correlations, leading to coding artifacts that deteriorate the visual quality and can be bothersome to viewers. To address this issue, compressed images and video sequences require enhancement before being presented on display devices. Traditional approaches to quality enhancement typically concentrate on enhancing individual frames, often neglecting to ensure temporal consistency across frames.

B. Lucas-Kanade's method- The Lucas-Kanade algorithm is a sparse optical flow method designed to estimate the motion of a selected subset of pixels within an image or video frame. It operates under the assumption that neighboring pixels exhibit similar motion patterns and employs a least-squares technique to compute the motion parameters**.**

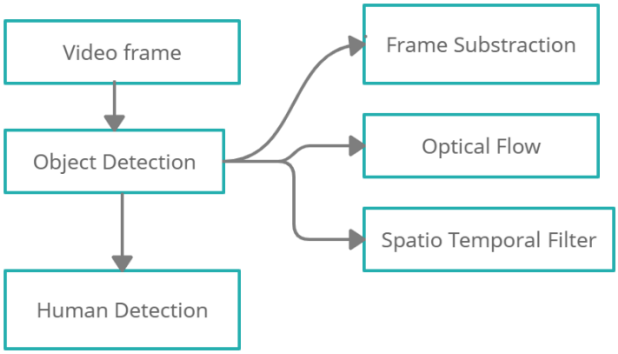

Figure2: Flow for implementing Human Detection

[II] Human Tracking- Tracking means to track the object in each time frame and to locate the same object in different frames to detect its motion. For tracking motion, the main categories considered are-

A. Point Tracking: Point tracking is a widely-used method in computer vision and object tracking, where specific points within an object or scene are identified and tracked across successive video frames. This method is divided into two primary categories: deterministic and statistical methods.

Deterministic point tracking relies on the precise location of specific points within an object to follow its motion through frames. These points could be corners, edges, or other distinctive features within the object.
Statistical point tracking, on the other hand, employs probabilistic models to estimate the most likely trajectory of the tracked points. Methods like the Kalman filter and particle filter fall under this category.

B. Kernel Tracking: Kernel tracking is a method for estimating the motion of an object by analyzing the changes in pixel values or intensity within a defined region (the kernel) as it moves from one frame to the next. This approach allows for the computation of object motion by observing how the pixel values evolve over time. Kernel tracking can be particularly useful when dealing with objects that do not have easily identifiable points or features, as it relies on the overall pixel-level information to estimate motion.

C. Silhouette Tracking: Silhouette tracking is a method for tracking objects in video frames by considering their geometric shapes or silhouettes. This approach involves segmenting the object from the background based on its outline or silhouette and then tracking this silhouette across frames.

Among all the tracking methods mentioned, the silhouette tracking method stands out for its ability to

provide precise and reliable results when applied to appropriate scenarios.

[III] Anomaly Detection and Behavior Understanding- This can be done by detecting the normal and abnormal behavior of people, using pattern and action detection.
Crowd Dynamics is a study of crowd, which investigates the possibility from where the crowd can generate and its actions, mobility, and behavior. To gather all these information, some critical parameters such as its size, direction of motion, the range of organized/violent crowd are taken into consideration. The speed of crowd motion can point towards some emergency. [3][4]
The initial step involves image processing, which encompasses the identification of stationary crowd elements. Subsequently, the estimation of crowd density is performed, with techniques such as edge detection, optical density estimation, and geometric distortion analysis being employed. It's worth noting that these methods, including others, are susceptible to a phenomenon known as the near-far effect, where individuals closer to the camera appear larger and occupy more space compared to individuals of the same size positioned farther from the camera.

3. PREVIOUS WORK

Detection of crowd density can be done by detecting the count of public and identifying the face and position at the same time in video. It can be quite tedious as images and video frame contains faces with different color, pose, expression and position of human. Adding on to it, various light conditions (illumination), different angles and orientation of camera add on to the difficulty level.
There are three approaches used here to estimate the correct count and density. One of the common approaches is by detecting only one person and their location simultaneously in the frame to increase the accuracy. In other approaches, Haar filter are used to capture the difference in different video frames, Haar Wavelet Transform (HWT) used to extract head feature and Support Vector Machine (SVM) to detect the head.

3.1 Haar Filter- Haar functions have a historical legacy dating back to 1910 when they were introduced by the Hungarian mathematician Alfred Haar [5]. Over time, various formulations of Haar functions, along with diverse representations [6], as well as multiple adjustments [7, 8, 9], have been published and implemented. One notable and highly regarded modification is the lifting scheme [10, 11, 12].

These transformations have found application in various domains, including spectral techniques for multiple-valued logic, image coding, and edge extraction, among others.

3.2 Support Vector Machine (SVM): To implement SVM a video is converted into frames and upon which preprocessing is applied. which is often termed as Median Filter. It then does Historical Equivalence which change original image pixel gray value if number of pixels in the image gray value to widen, while the number of pixels in a small level reduction, the image is converted into form of histogram. Segmentation is done by breaking frames into various segments (set of pixels). The object detection is performed and SVM classification is done which classifies crowd into Abnormal or normal crowd [14].

A Regression-based approach is employed to map features for counting objects in images. Methods such as Histogram of Oriented Gradients (HOG) and Gray Level Occurrence Matrices (GLCM) have been utilized to enhance the accuracy of results. With the introduction of CNN and deep learning methodologies for crowd detection, it becomes possible to obtain an end-to-end count directly from video frames. The Multi-Scale Convolutional Neural Network (MSCNN) takes the entire image as input, in contrast to processing small patches. The Density-Based technique establishes a linear correlation between neighborhood direction capabilities and their corresponding object density maps.

3.3 Histogram of Oriented Gradients- This algorithm helps to fine the count of the person in each frame, and measuring the difference in the count of subsequent frames can lead to accurate results and talk about the increase in the density of the crowd. This method generates a histogram in the end to show the appearance of the human detected in the video file (Figure2).

3.4 CNN- The work encompasses two distinct approaches: one involves processing image patches, while the other takes the entire image into account. The utilization of Multi-Column CNN enables the handling of images with varying sizes or resolutions, facilitating the capture of details even in low-resolution images and those captured from different angles.

Crowd analysis poses several challenges, including high clutter, variations in contrast, poor resolution, and same and other frame differences. To address these challenges, the previously mentioned algorithms have been introduced.

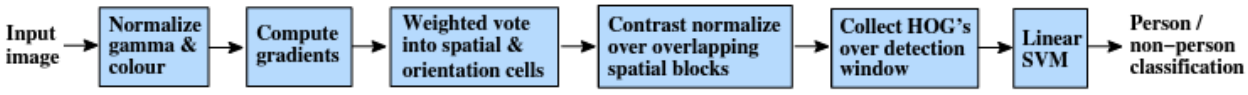

Figure3: Illustrates an overview of our feature extraction and object detection process. [16].

Out of the three algorithms, Regression employs Deep CNN, which offers more robust and precise results. Several strategies have been devised to mitigate side effects in images, leading to improved accuracy. Techniques such as incorporating Haar filters and applying layers on grayscale images have demonstrated enhanced performance in face detection.

4. REAL-TIME SURVEILLANCE SYSTEM

The recent update made on a Real-time surveillance system using CNN is proposed in the paper [17]. This method includes removing the background of images and decreasing the unneeded pixels from the image so that image can be used to evaluate a crowd activity whether it is normal or abnormal. It uses VGG16 as reference to reduce time complexity while maintaining good detection performance.
There are weighted layers and only fourteen of the architecture's total 19 layers of which eight convolutional, five Max Pooling and three Dense are weight layers. Architecture's input tensor has a size of 64, 64 and three RGB channels.

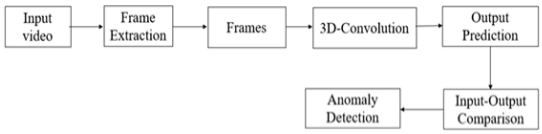

Figure4: Flowchart of steps performed in CNN [17]

Another approach includes Harr Cascades detection, it works on principle that first using a dataset we train a model and then real time video feed is given to classify the data. It includes three major principles

Haar Cascade Classifier- Detection the use of Haar function-primarily based cascade classifiers is a good method delivered with the aid of Paul Viola and Michael Jones in their paper titled 'speedy object Detection the use of a Boosted Cascade of easy functions' in 2001.This approach is grounded in machine learning, where a cascade function is trained using a substantial dataset of positive and negative images. Subsequently, this trained model is employed to detect objects in other images [18].

Figure5: Gray Scaled image [18]

Adaboost Algorithm- AdaBoost, short for Adaptive Boosting, is an ensemble boosting classifier introduced in 1996. This technique employs an iterative process where multiple classifiers, which individually exhibit weak performance, are amalgamated to create a robust single classifier, thereby enhancing efficiency.

Optical Flow Algorithm- The optical flow algorithm is rooted in the analysis of how images in two consecutive frames create a motion pattern as a result of either object movement or camera motion. In this context, each pattern represents a 2D displacement vector field, with the arrow vector indicating the direction of object movement between the two frames.

5. RESULT AND DISCUSSON

A detection-based totally approach has demonstrated success, in scenarios with low-density crowds. This method employs Haar Wavelet remodel (HWT) and assist Vector machine (SVM) for the type of heads and their respective components. additionally, a Kanade Lucas Tomasi (KLT) tracker is utilized to extract a hard and fast of low-level functions, aiding within the identity of moving objects inside video frames.

The regression-based approach encompasses various techniques for capturing both local and global scene properties. These techniques include Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). Following the extraction of local and global features, neural networks are applied to further analyze the data.

Model evaluation- often involves using accuracy as a metric to assess the model's performance across all classes. Accuracy is obtained by dividing the total number of right predictions made upon the total number of predictions made. The total number of predictions made is nothing but sum of true positives and negatives.

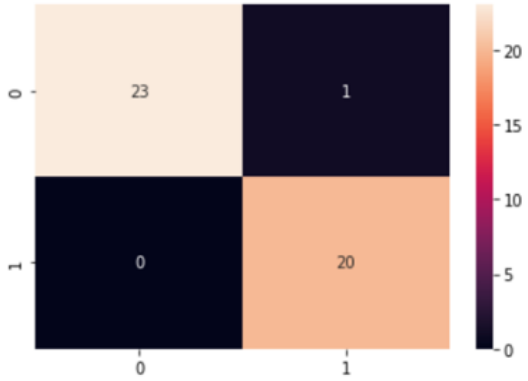

Figure6: Confusion matrix of CNN approach [17]

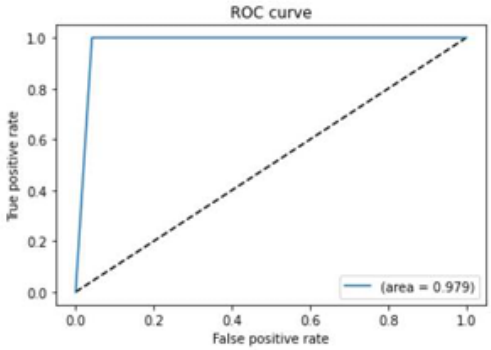

Figure7: ROC curve for Indoor Evaluation [17]

Density based algorithms makes use of Random Forest Regression from a couple of image patches for vote casting densities of a couple of goal options.

6. CONCLUSION

Over the past few decades, Closed-Circuit Television (CCTV) surveillance has become an integral component of monitoring crowded public areas. This proliferation of surveillance cameras has generated a massive volume of video data, necessitating automated solutions for its efficient analysis and management. The primary motivations for automated video surveillance include:

[I] Safety Control- Video surveillance cameras for safety of people at public places such as malls, stadium to detect and monitor the crowd behavior and any abnormal activity.
[II] Disaster management- Monitor the situation at crowded places to prevent the risk of any situation of havoc or stampede.
[III] Visual Surveillance- Automated surveillance systems are capable of alerting authorities to unusual activities in public places, enabling a rapid response to potential security threats or suspicious behavior.

To mitigate the potential risks associated with crowded places, it becomes imperative to implement automatic detection of abnormal and critical situations. Crowd density detection stands as one of the most intricate challenges within the realms of CV and field of ML. The vital approaches have emerged for this purpose: Detection-Based, Regression-Based, and Density-Based Approaches. In this context, leveraging deep learning models, particularly Deep Convolutional Neural Networks (CNN), proves to be a powerful tool for detection. Unlike traditional methods that operate on image patches, deep CNNs process the entire image through multiple layers, offering the potential for improved accuracy.
However, achieving high accuracy in crowd density detection remains a complex endeavor. While CNNs hold promise, factors such as orientation and illumination introduce complexities that necessitate careful consideration and further research.

There are still some issues that need to be solved to improve the results of automatic video surveillance:

A. Tracking the Detection Framework- Developing robust tracking mechanisms to follow individuals across frames is essential for maintaining accurate crowd density estimations, particularly in dynamic environments.
B. Analyzing Crowd Behavior- Understanding and predicting crowd behavior, including identifying abnormal or potentially dangerous patterns, is an ongoing research area critical to enhancing surveillance systems' effectiveness.
C. Real-Time Processing and Generalization- Achieving real-time processing capabilities and ensuring that surveillance models can generalize well across diverse environments and conditions are essential goals for improving system reliability.

7. REFERENCES

[1] Lamba, Sonu & Nain, Neeta. (2016). A Literature Review on Crowd Scene Analysis and Monitoring. International Journal of Urban Design for Ubiquitous Computing, 4, 9-20. DOI: 10.21742/ijuduc.2016.4.2.02.
[2] Chaudhari, Mayur & Ghotkar, Archana. (2018). A Study on Crowd Detection and Density Analysis for Safety Control. International Journal of Computer Sciences and Engineering, 6, 424-428. DOI: 10.26438/ijcse/v6i4.424428.
[3] Davies, A.C., J.H. Yin, and S.A. Velastin. Crowd monitoring using image processing. Electronics & Communication Engineering Journal, 1995, 7(1), 37-47.
[4] Marana, A., et al. Estimation of crowd density using image processing. Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium on, 1997. IET.

[5] Haar, A. "Zur Theorie der orthogonalen Funktionensysteme." Mathematische Annalen, 69, 331–371.
[6] Zeng L., Jansen C. P., Marsch S., Unser M., Hunziker R. "Four–Dimensional Wavelet Compression of Arbitrarily Sized Echocardiographic Data." IEEE Transactions on Medical Imaging, 21(9), 1179–1188.
[7] Claypoole R., Davis G., Sweldens W., Baraniuk R. "Adaptive Wavelet Transforms for Image Coding." Asilomar Conference on Signals, Systems and Computers.
[8] Munoz A., Ertle R., Unser M. "Continuous wavelet transform with arbitrary scales and O(N) complexity." Signal Processing, 82, 749–757.
[9] Porwik P., Lisowska A. "The New Graphic Description of the Haar Wavelet Transform." Lecture Notes in Computer Science, Springer–Verlag, Berlin, Heidelberg, New York, 3039, 1–8.
[10] Porwik, Piotr & Lisowska, Agnieszka. (2004). "The Haar-wavelet transform in digital image processing: its status and achievements."
[11] Daubechies I. "Recent results in wavelet applications." Journal of Electronic Imaging, 7(4), 719–724.
[12] Daubechies L., Sweldens W. "Factoring wavelet transforms into lifting steps." J. Fourier Anal. Appl., 4(3), 247–269.
[13] Davis G., Strela V., Turcajova R. "Multivwavelet Construction via the Lifting Scheme." Wavelet Analysis and Multiresolution Methods, T. X. He (editor), Lecture Notes in Pure and Applied Mathematics, Marcel Dekker.
[14] Patel, Prof & Patel, Ravi & Student. (2019). "SUPPORT VECTOR MACHINE (SVM) BASED ABNORMAL CROWD ACTIVITY DETECTION."
[15] Surasak, Thattapon & Takahiro, Ito & Cheng, Cheng-hsuan & Wang, Chi-en & Sheng, Pao-you. (2018). "Histogram of Oriented Gradients for Human Detection in Video."
[16] Navneet Dalal, Bill Triggs. "Histograms of Oriented Gradients for Human Detection." International Conference on Computer Vision & Pattern Recognition (CVPR '05), Jun 2005, San Diego, United States. pp. 886--893, DOI: 10.1109/CVPR.2005.177.
[17] Amina P, Dr. Binu L S. "Real-Time Crowd Analysis and Anomaly Detection," 08 September 2022, PREPRINT (Version 1) available at Research Square.
[18] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Mitsubishi Electric Research Labs, Cambridge, 2001, IEEE.
[19] Hentschel T., 1993, "Image Processing Techniques for the Estimation of Features of Crowd Behaviour in Urban Environments," MSc. Dissertation, King's College London, UK.
[20] Polus A., Schofer J.L. and Ushpiz A., 1983, "Pedestrian Flow and Level of Service," J. Transportation Engineering, 109, 46-56.
[21] Brown R.G. and Hwang P.Y.C., 1992, "Introduction to Random Signal Analysis and Kalman Filtering," Wiley, 2nd ed.