International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

# CROWD MONITOR : ADVANCED SYSTEM FOR HUMAN DETECTION AND COUNTING IN SURVEILLANCE

***Abin James*[1]**

Department of computer science and IT, Student
Jain (Deemed To Be) University, Bengaluru, India
Email address: abinjames2002@gmail.com

ABSTRACT:

This project's Human Detection and Counting System uses machine learning algorithms and computer vision techniques to identify and count people in pictures and videos. The system is intended to handle multiple real-world applications, including traffic analysis, security monitoring, and crowd management. For effective implementation, the project makes use of the Python programming language and well-known libraries like TensorFlow and OpenCV. The method consists of several steps, such as feature extraction, classification, and preprocessing. To improve the quality of the data, preprocessing techniques like scaling, normalization, and noise reduction are first applied to the raw photos or video frames. In order to extract pertinent characteristics from the input data, feature extraction techniques are utilized. To determine if an object is human or not, these attributes are then fed into a trained machine learning model, which is frequently built using deep learning architectures like convolutional neural networks (CNNs).Because of its strong construction, the system can withstand changes in position, occlusions, and lighting. It is also scalable to meet the demands of real-time processing and massive datasets.F1-score, accuracy, precision, recall, and other quantitative measures are used to assess how effective the suggested approach is. Qualitative evaluations are also carried out on real-world datasets to confirm the system's practical usefulness. All things considered, the Human Detection and Counting System described in this project offers a dependable and effective way to automate the process of human detection and counting in a variety of settings, advancing both public safety and surveillance technology

## 1. Introduction :

The COVID-19 epidemic has had a profound effect on daily living, organizations, and international interactions, drastically altering society. Healthcare professionals and researchers are leading the charge in the fight against the virus, which has brought attention to the significance of both human action and technology—particularly machine learning (ML). By utilizing AI technology, we can keep an eye on human activities and detect it in order to maintain social distance and stop the COVID-19 virus from spreading. The goal of this research is to solve the problem of precisely counting people in complicated scenarios captured in low-resolution photos. This is important for a variety of surveillance applications, including crowd control and security control.

This technology can help with crowd monitoring, dispersing the risk of viral outbreaks, and monitoring public areas by devising a reliable technique for counting crowds and identifying individuals. It also provides organizations with insightful information that helps them to determine peak traffic hours, optimize personnel levels, and assess worker performance across several locations.

Strong and effective human detection and counting systems are becoming more and more important in today's world, as shown in the security, retail, transportation, and urban planning sectors, among other areas. Conventional techniques for human detection and counting sometimes depend on laborious manual labor or overly basic algorithms, both of which are prone to mistakes and inefficiencies—especially in dynamic contexts with changing lighting and complicated backgrounds. In order to tackle these issues, this project presents a thorough Human Detection and Counting System that makes use of cutting-edge Python machine learning and computer vision techniques. The system attempts to automate the process of reliably and accurately detecting and counting humans in photos and videos by utilizing computer vision and machine learning. This project's main goal is to create a flexible, scalable system that can reliably identify and count people in a variety of settings, from congested city streets to regulated interior spaces. Beyond simple detection, the system's capabilities aims to offer real-time analysis and insights for applications including traffic monitoring, crowd control, and security surveillance. Preprocessing approaches to improve image quality, feature extraction techniques to retrieve pertinent information, and machine learning models—particularly convolutional neural networks (CNNs)—for reliable classification are important parts of the proposed system. Through the seamless integration of these components, even under difficult settings, the system is able to identify humans from other objects and background clutter. In addition, the research highlights how crucial it is to validate and assess performance on real-world datasets in order to guarantee the system's dependability and practical applicability. The project seeks to show the effectiveness and efficiency of the proposed Human Detection and Counting System through extensive testing and comparison against known metrics.

Overall, this study offers a comprehensive response to the urgent demand for precise and dependable human identification and counting systems in a variety of domains, making a substantial contribution to the fields of computer vision and machine learning. The system has the ability to improve decision-making processes in various real-world settings, optimize resource allocation, and boost security through its creative methodology and practical implementation.

## 2. Literature Review :

[1] The purpose of this study is to suggest a procedure for calculating the population in different situations. It uses two main methods for foreground detection: the first is the mixture of Gaussians approach by Stauffer and Grimson, which represents each pixel as a weighted combination of Gaussian distributions and uses this to accomplish adaptive background subtraction. Second, the scene is approximately divided into layers that represent the image background using the pixel layering method. A Gaussian kernel is used in non-parametric kernel density estimation to determine the likelihood that a pixel is part of a layer.

[2] An Automated Grow Cut (AGC) segmentation technique is used by the system to effectively classify foreground regions based on labelled seed pixels obtained from the initial depth analysis. Compared to supervised segmentation techniques, this method allows the segmentation of human elements with less computing overhead. In order to obtain important distance information about the contents of the scene in relation to the camera, RGB-D cameras are used to collect both color and depth information in a picture. These camera sensors are inexpensive and easy to use. The system makes use of the Grow Cut segmentation technique and an unsupervised segmentation strategy based on depth data. This segmentation procedure is fast enough to accommodate intermediate frames, making it useful for applications like human tracking and activity recognition.

[3] This work introduces a novel method for precisely estimating the number of people and identifying each individual in complex scenarios using low-resolution photos. First, post-processing processes are applied to background-subtracted findings in order to count the number of people in a scene. After that, an Expectation-Maximization (EM) based method is presented to accurately locate people in these low-resolution scenarios. This approach uses a new cluster model to represent every individual in the picture. In addition, the number of people is used as a priori data to locate people using feature points. As a result, the approaches for locating and estimating the number of individuals are smoothly integrated, improving the approach's overall efficacy.

[4] Indoor counting technologies are essential for supporting rescue efforts because they give emergency responders critical information about how many people are leaving a building and where they are on each floor. Of these methods, Life Count presents a unique two-step neural network technique that is intended to accurately count individuals moving through a hallway. This revolutionary method—dubbed Life Count—uses a CSI-based human counting algorithm that is more accurate than any of the previously published methods. In contrast to other methods, Life Count counts numerous people accurately at once who cross the sensing region by using a single TX/RX link.

[5] Three important processes are preprocessing, human detection, and human tracking in the development of motion-based models and neural networks (CNN) for person detection in infrared video feeds. In order to improve the raw image quality, noise cancellation filters and histogram equalization are applied during the preprocessing step. The enhanced image is then subjected to CNN-based person detection, which makes use of its capacity to precisely identify human figures. Furthermore, to improve the overall detection speed, an adaptive tracking of detected individuals is made possible by the integration of a motion-based model that makes use of the KCF Tracker. To maximize computing efficiency, the architecture includes depth-wise separable convolution, a blend of depth-wise and pointwise convolutions. The detection rate was 5.1% higher with noise filters and histogram equalization applied during the preprocessing phase than with an unprocessed method. Additionally, applying AI-based human identification offers a practical method for precise people recognition in infrared video feeds, greatly resolving object localization difficulties.

## 3. Existing System :

Techniques for calculating the quantity of people and their locations typically function in tandem. This is accomplished by detection-based techniques, which recognize people in a given scene. But dealing with low-resolution camera images presents difficulties and frequently impedes accurate detection. Another major obstacle to background segmentation approaches is occlusion. Machine vision finds it challenging to detect items like humans because of their wide variety of appearances. Positively, localized techniques are at the core of recent advances in human detection. Although human counting and detection have advanced, there is still a need for an intuitive application that allows users to see these counts for themselves.

## 4. Proposed System :

In visual surveillance, people counting and human detection are crucial responsibilities. However, current methods frequently have drawbacks, such as the need for high image quality, simple backgrounds, and mobility. We provide a comprehensive system that can count the number of individuals and detect humans in order to overcome these issues. The device uses camera footage to record real-time video, which is then separated into individual frames to precisely count the number of people. A different programmed then visualizes this count, giving insights into the population density of various regions. Furthermore, the count provides a foundation for individual location utilizing feature points, unifying the techniques for both population estimation and individual location. The purpose of this suggested approach is to reduce crowding in public areas like stores, expedite procedures, and save time.

## 5. Algorithm :

- The major input source of the system is a CCTV or webcam, which is used to record live video footage.
- Subsequently, the obtained video stream undergoes processing to isolate discrete visual frames, each signifying a unique instant in time.
- Based on their individual timestamps, these image frames are arranged in a sequential manner to create a chronological series of visual data.
- After that, each image frame is converted into a N × N array grid to make item detection and analysis easier.
- This grid layout uses an item detection algorithm to determine if an object is present or absent by predicting its existence within each grid cell.
- The boundaries of recognized objects are then properly defined using bounding box regression algorithms, improving the accuracy of localization.
- A non-maximum suppression technique is used to systematically remove weak or overlapping predictions in order to improve the detection results and get rid of redundant predictions.
- After suppression, a filtering procedure isolates and differentiates human objects from other detected objects by explicitly targeting them within the discovered entities.
- Real-time monitoring and analysis of human presence in the video feed is made possible by the system, which keeps track of the number of individuals found in each frame and stores this data for user access through a special Android application.

## 6. System Diagram :

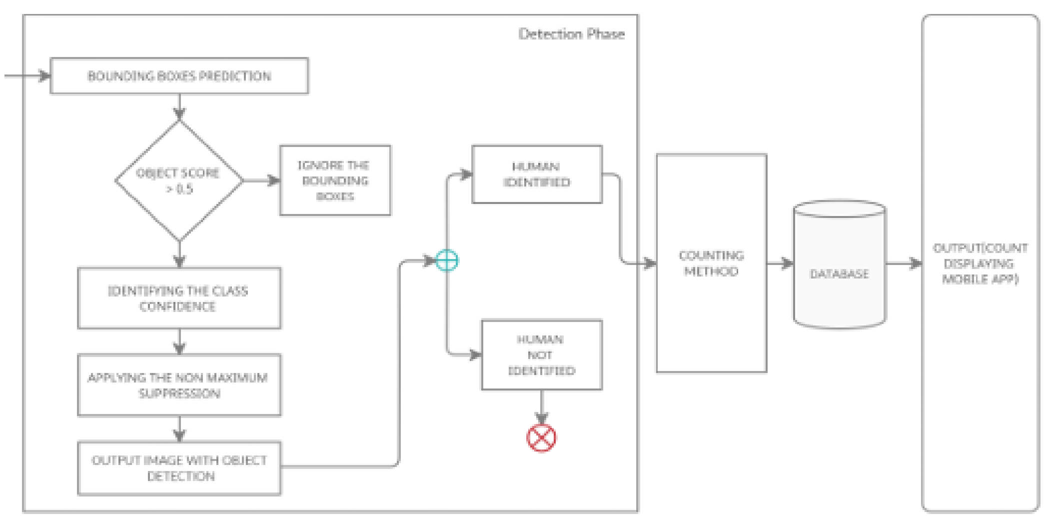


**Figure-1: system diagram**

### 6.1 Image Acquisition :

The first phase of our system is image acquisition, which is the basic procedure of taking a picture from a source, usually a piece of hardware, in order to make further processing easier. This is a crucial stage in the workflow since it forms the basis for all other image-related actions. The process of acquiring images is essential because it makes it possible to extract the visual data needed for analysis and interpretation. By using cameras or scanners, pictures are captured in a way that maintains the integrity of all relevant elements. This guarantees that the obtained images faithfully capture the desired content, offering a solid foundation for further image processing operations.

### 6.2 Frame Separation:

Real-time video footage recorded by Closed-Circuit Television (CCTV) or comparable video capturing equipment is used as input in our suggested solution. However, the video feed must be divided into individual frames or images in order to process and identify each individual. After being gathered, these photos are arranged into an image sequence in order to be processed and examined further.

### 6.3 Image separation:

A picture sequence is a collection of successive still photos that show different animation frames. To maintain their sequential order, these photos are usually kept in a single folder and labelled with file name extensions. Videos or image sequences are very useful in real-world applications; they are frequently used for motion capture, data processing, and visual storytelling.

## 7. YOLO Algorithm :

The technique known as "You Only Look Once," or YOLO, uses neural networks to provide real-time object recognition. YOLO, which is well-known for its speed and accuracy, is used extensively in many different fields, such as the identification of people, animals, parking meters, and traffic signals. Based on the idea of instantly identifying and detecting many items in an image, YOLO approaches object identification as a regression issue and returns classification probabilities for objects that are found.

The unique characteristic of this approach is that it uses convolutional neural networks (CNN) to achieve real-time object detection. To identify objects, it only has to make one forward pass through the neural network. With this single-pass method, predictions may be made for the whole image in a single computation run. The CNN uses methods including residual blocks, bounding box regression, and intersection over union (IOU) to improve detection efficiency and accuracy while simultaneously predicting class probabilities and bounding boxes

$$y = (P_c, b_x, b_y, b_h, b_w, c)$$

**Figure-2: Yolo algorithm equation**

## 8. Database :

The controlled NoSQL document database Firebase Realtime Database is intended for online and mobile application development. It is presently in beta and provides effective worldwide application data synchronization and storage. The flow that is normally followed when storing data in Firebase Realtime Database is shown in the accompanying figure. The JSON format is mostly followed by the database structure. Details on every organization in this project, including name, address, image, and variable count, are kept in the primary directory called "customers." Every organization in this directory has a special identification assigned to it so that it may efficiently organize and retrieve its particular data.

## 9. Conclusion :

The next stage after building the model is to test and evaluate its functionality. Owing to its training on the COCO dataset, the model is anticipated to identify individuals in both picture and video formats. Videos and photos were used as inputs during testing, and the output that was produced is shown in the accompanying figure. The quality of the input photos or video frames is one of the many criteria that affects how well the model performs. Digital colour photographs perform well for detection, while black-and-white images perform poorly. Furthermore, the camera's placement is critical since obstructions like electric posts or trees might impede its field of vision and reduce the precision of the detection.

Beyond human detection, the suggested method allows counts to be detected and displayed through a mobile application. This real-time data helps users make decisions about when to visit places like public distribution systems, banks, and supermarkets by enabling them to evaluate the current volume of people at different establishments. Although there is still space for improvement in the current model's accuracy and latency, next improvements might incorporate motion detection features to identify moving objects. Additionally, YOLO v3's performance might be improved by adding straightforward training heuristics and making architectural changes. Utilizing GPU acceleration methods and CUDA packages could improve system performance by increasing processing speed and efficiency.


**Figure-3: result of project**

REFERENCES :

[1] D. Fehr, R. Sivalingam, V. Morellas, N. Papanikolopoulos, O. Lotfallah, and Y. Park, "Counting people in groups," 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2009, pp. 152–157, 2009, doi: 10.1109/AVSS.2009.55.

[2] H. Fernando, I. Perera, and C. De Silva, "Real-time Human Detection and Tracking in Infrared Video Feed," MERCon 2019 - Proceedings, 5th Int. Multidiscip. Moratuwa Eng. Res. Conf., pp. 111–116, 2019, DOI: 10.1109/MERCon.2019.8818862.

[3] Y. L. Hou and G. K. H. Pang, "People counting and human detection in a challenging situation," IEEE Trans. Syst. Man, Cybern. Part ASystems Humans, vol. 41, no. 1, pp. 24–33, 2011, DOI: 10.1109/TSMCA.2010.2064299.

[4] D. Konings and F. Alam, "LifeCount: A Device-free CSI-based Human Counting Solution for Emergency Building Evacuations," 2020 IEEE Sensors Appl. Symp. SAS 2020 - Proc., pp. 1–5, 2020, DOI: 10.1109/SAS48726.2020.9220032.

[5] P. Lestari and H. P. Schade, "Human Detection from RGB Depth Image using Active contour and Grow-cut Segmentation," 2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019, pp. 70–75, 2019, DOI: 10.1109/IC3INA48034.2019.8949571.

[6] B. Li, J. Zhang, Z. Zhang, and Y. Xu, "A people counting method based on head detection and tracking," Proc. 2014 Int. Conf. Smart Comput. SMARTCOMP 2014, pp. 136–141, 2014, doi: 10.1109/SMARTCOMP.2014.7043851.