Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 15, Number 1 (2023), pp. 33-46

https://dx.doi.org/10.37624/JCSA/15.1.2023.33-46

People Counting and Tracking System in Real-Time Using Deep Learning Techniques

Dr. O. Obulesu[1*], A. Hanshika[2], A. Lahari[3], P. Arunodaya[4], R. Ashwini[5]

[1,2,3,4,5]Department of CSE, G. Narayanamma Institute of Technology & Science, Hyderabad, Telangana, India
**Email: obulesh194@gmail.com*

Abstract

The aim of project is to automatically estimate the number of people at indoor and outdoor places. People counting systems can be used in retail environment such as determining conversion ratio, advertising and promotional evaluation. This system can be used for transportation management system and video surveillance. The number of customers is indispensable data for management and decision making in public places like large-scale markets, shopping centers, airports, stations, museums, laboratories, classrooms, cafeteria etc. In this system, firstly extract the frames from the video, then draw a desired reference line on the input frame, detect the people using MobileNet-SSD object detection model, mark the centroid on the detected person, track the movement of that marked centroid and calculate the direction of centroid movement whether it is moving upwards or downwards. If the centroid movement is downward direction, then we increment in counter, else if the centroid movement is upward direction, then we increment out counter. People counting is essential for retailers of any size, but it's especially important for small businesses that don't have the benefit of assaying data from multitudinous locales when making pivotal opinions. When used intelligently, people counting can shape businesses in multitudinous ways other than just furnishing information on nethermost business.

Keywords: Deep Learning, MobileNet-SSD, Centroid Tracking

1. Introduction

The analysis of crowds is a popular research topic that has gained attention due to its various practical applications. Automated people counting systems are designed to estimate the number of individuals present in both indoor and outdoor environments.

These counting technologies serve as the foundation for a range of advanced solutions, such as retail analytics, queue management, and space utilization. The use of real-time information on people flow is especially beneficial for applications related to crowd management and security, including the management of pedestrian traffic and visitor flow evaluations. It is widely acknowledged that tracking and monitoring the movement of people in an area is of significant importance. People counting and tracking systems have been used for many years in various applications. For example, in the early 20th century, manual people counting systems were used in retail stores to track the number of customers. Later, in the 1960s and 70s, mechanical and electronic systems were developed to automate people counting.

In the 1990s, video-based systems were introduced, and since then, there have been numerous technological advancements in computer vision and artificial Intelligence that have greatly improved the accuracy and efficiency of these systems. In today's world, it is crucial for companies to consider potential solutions that can ensure the safety of both staff and customers. This is especially important as new legislation is being introduced worldwide to regulate occupancy and social distancing. The importance of people has never been greater, and it is now possible to accurately count visitors using technology. Sensors connected to software can provide real-time occupancy data, which helps determine whether a location can safely accommodate more people or not. By using this technology, overcrowding can be prevented before it becomes a serious issue, ensuring that a location does not exceed its capacity. This is essential in keeping both customers and employees safe. Collecting data on the efficiency of a shop's operations is essential for maximizing a company's potential. By monitoring customer satisfaction, companies can implement ideas to enhance the consumer experience, which can lead to increased revenues and customer loyalty. People counters are a practical solution for counting and monitoring the flow of visitors in and out of a facility. The information gathered by these counters can be segmented and used to determine visitor arrival and departure times, as well as the level of engagement customers have with products and services. People counting technology can also reveal where visitors spend the most time in a shop and which products are most popular.

This information is valuable for companies seeking to optimize their marketing efforts and increase sales. In the retail industry, people counting technology is used to determine the precise number of consumers who visit a shop. People counting technology can provide valuable insights to executives and organizers of events and exhibits. By monitoring foot traffic and fluctuations throughout the day and week, executives can design effective marketing programs and plan for the future. For event organizers, people counting technology can be used to report the total number of participants and identify the event's busiest hours. This information allows management to assign staff members more effectively, based on the number of people attending the event. These insights can help ensure that events and exhibits run smoothly and provide a positive experience for attendees. People counting technology has been around for several decades, and its development has been shaped by various technological advances and market demands. The evolution of people counting technology has been quite significant, from manual tallying methods to modern sensor-based systems, and

has been driven by the need for more accurate and realtime data on human behaviour. The key drivers behind the adoption of people counting technology have been the increasing demand for data-driven decision-making, the need for efficient resource allocation, and the desire to enhance customer and patient experience.

There are numerous sorts of human beings counting technology, every with its personal benefits and disadvantages. Thermal imaging, infrared sensors, video analytics, and RFID (radio-frequency identification) are some of the most common types of people counting technology. Thermal imaging is based on detecting body heat, while infrared sensors use beams of light to detect human presence. Video analytics use cameras and software to analyse video footage, while RFID relies on radio signals to detect the presence of a tag or card. Each of these technologies has its own unique features and is suitable for different applications. People counting technology generates large amounts of data, which must be collected, processed and analysed to derive insights. There are different methods of data collection, such as cloud-based systems and edge computing. Cloud-based systems store data on remote servers, while edge computing processes data on local devices, such as sensors or cameras. The tools and techniques used for data analysis include machine learning and predictive analytics, which help to identify patterns and trends in the data and make predictions about future behaviour accuracy and reliability of people counting technology are crucial for its usefulness and adoption. Several factors influence accuracy and reliability, such as sensor placement, lighting conditions, and occlusions. The methods used to validate the accuracy of people counting systems include manual counting, video verification, and statistical analysis. Validation is essential to ensure that the data generated by people counting technology is accurate and reliable, and can be used for decision making.

2. Related Works

The task of people counting and tracking is of great importance in many fields, including security, retail, transportation, and crowd management. Over the years, various methods have been proposed for people counting and tracking, ranging from traditional approaches such as video analysis and sensor networks to more advanced techniques based on deep learning. One of the earliest and most widely used approaches for people counting and tracking is video analysis. Video analysis techniques involve processing video data to detect and track people in real-time. These techniques can be classified into two categories: frame-based and pixel-based. Frame-based techniques, such as background subtraction, template matching, and optical flow, analyse each frame of the video independently to detect and track people. Pixel-based techniques, such as blob analysis and contour tracking, analyse the movement of pixels over time to detect and track people. While these techniques can be effective in simple scenarios, they are often limited by issues such as occlusions, changes in lighting conditions, and the presence of clutter in the environment. To overcome the limitations of traditional video analysis techniques, researchers have proposed various deep learning-based approaches for people counting and tracking. Deep learning techniques are based on the use of neural networks to automatically learn features from large amounts of data. These techniques have shown promising results in a variety of computer vision tasks,

including object detection, recognition, and tracking. Some of the most commonly used deep learning techniques for people counting and tracking include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks. One of the most popular deep learning techniques for object detection and recognition is Convolutional Neural Networks (CNNs). CNNs have been successfully applied to various computer vision tasks, including object detection and tracking.

D. Zhang et al. [1] reviewed the existing techniques for people counting and crowd density estimation, including traditional methods and recent advances in deep learningbased approaches and provided a comprehensive overview of the various techniques, their strengths and limitations, and their applications in different scenarios. He further highlights the importance of accurate people counting and tracking in crowd management.

S. L. Choi et al. [2] developed a multi-camera people counting system using deep learningbased object detection and tracking techniques. The system uses a combination of CNNs and LSTMs to detect and track people across multiple cameras in real-time. The system is evaluated on a public dataset and achieves high real-time performance.

K. Raja et al. [3] provided an overview of various deep learning techniques for object detection and recognition, including their strengths and limitations. He discussed about some popular techniques such as Faster R-CNN, YOLO, and SSD, and compares their performance on different datasets, but does not focus specifically on people counting and tracking, it provides a valuable overview of the state-of-the-art deep learning techniques in the field of computer vision. Another deep learning technique that has been used for people counting and tracking is Recurrent Neural Networks (RNNs). RNNs are designed to handle sequential data, making them well-suited for tasks such as tracking objects over time.

J. Guo et al. [4] propose a pedestrian detection system using RNNs. The system is designed to detect pedestrians in video sequences and track them over time. An analysis of the existing works shows the effectiveness of deep learning techniques for people counting and tracking. These techniques have been applied successfully in various scenarios, including object detection, recognition, and tracking. The use of multiple cameras and the integration of different deep learning techniques have also been investigated. While these techniques have shown promising results, there are still challenges to be addressed, such as occlusions, changes in lighting conditions, and the presence of clutter in the environment. Future research in this area should focus on addressing these challenges and developing more robust people counting and tracking systems.

Joseph Redmon et al. [5] proposed a real-time object detection system using CNNs. The system is called YOLO (You Only Look Once) and is designed to detect objects in real-time. The system is evaluated on several datasets and achieves high performance in real-time.

3. Proposed Methodology

The objective of this project is to develop a system capable of tracking and keeping a

record of the number of individuals present in a given location. The system can track the movement of people entering and leaving a building and stores this information. If the number of people exceeds a specific threshold, an alert message is sent to the user. A webcam is used as the video input source. The FPS throughput rate is determined, and the frames are converted to RGB and resized for image processing and computer vision using OpenCV. The OpenCV library is also used to perform deep neural network inference, to show output frames on the screen, and to open and write video files. The proposed system employs Object Detection and Object Tracking in two phases to increase accuracy. For Object Detection, Mobile-Net Single Shot Detector (SSD) is used, which is a pre-trained deep learning model. We only run it once every N frames due to its computational cost. To track detected objects, we use a combination of Correlations filters and centroid tracking algorithm.

To track objects, the coordinates of the bounding boxes are used to determine the centre or centroid. We then calculate the Euclidean distance between existing and new centroids and assign unique object IDs. Objects that enter the field are registered, and those that leave are deregistered. The Dlib library is used for implementing object tracking. OpenCV and Python are utilized for the implementation and prediction of crowd count.

This system has various practical applications, such as monitoring the number of people in public spaces like shopping centres, museums, and libraries, and alerting authorities when the number of people exceeds the permitted threshold. It can also be used for security purposes in restricted areas where only a limited number of people are allowed to enter. Furthermore, the system could be used to monitor and manage large crowds at events, concerts, or festivals, ensuring that the capacity of the venue is not exceeded.

3.1 Description of MobileNet-SSD

MobileNet SSD is an item detection version that computes the output bounding field and item elegance from the enter image. This Single Shot Detector (SSD) item detection version makes use of MobileNet as a spine and might reap rapid item detection optimized for cell devices. The architecture is shown in Fig 1 MobileNet SSD layered Architecture.

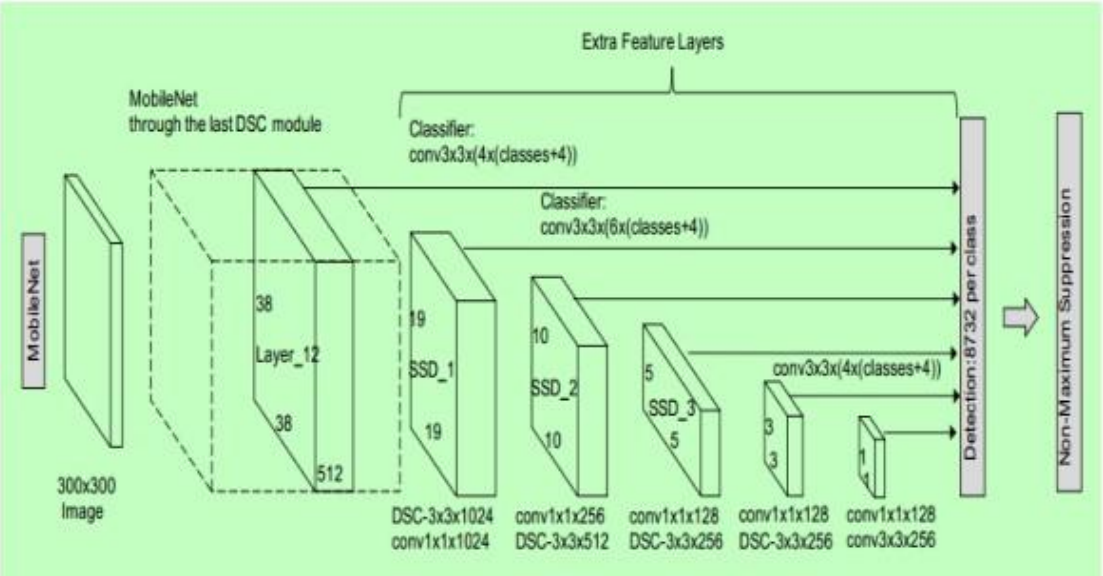

Figure 1: MobileNet SSD layered Architecture

This model is optimized for mobile devices and provides fast and accurate object detection. The architecture of MobileNet SSD consists of two main components: MobileNet and Single Shot Detector (SSD). MobileNet is a lightweight deep neural network architecture that is designed for mobile devices with limited computational resources. It is built upon depth-wise separable convolutions, which decompose a standard convolution into a depth-wise convolution and a point-wise convolution. The depth-wise convolution applies a single filter to each input channel, while the point-wise convolution applies a 1x1 filter to each output channel of the depth-wise convolution. This approach significantly reduces the number of parameters in the network, resulting in a smaller and more efficient model.

MobileNet is used as a feature extractor in the MobileNet SSD architecture. The input image is passed through several convolutional layers, with the depth-wise separable convolutions applied at each layer. The output of the last convolutional layer is a feature map, which contains high-level features that are useful for object detection. The SSD component of the architecture takes the feature map generated by MobileNet and performs object detection on it. The SSD approach is a single-shot detector that predicts the object class and location in a single forward pass of the network. The SSD approach involves dividing the input image into a grid of cells, with each cell responsible for predicting the presence of objects within it.

For each grid cell, the SSD predicts a set of bounding boxes, each with a different aspect ratio and size. These bounding boxes are defined as offsets from the coordinates of the grid cell. The SSD also predicts the confidence score for each bounding box, which represents the likelihood that the object is present in that box. The confidence score is calculated based on the overlap between the predicted bounding box and the ground truth bounding box. The SSD also predicts the class of the object in each bounding box.

The class prediction is based on a set of learned features that are specific to each object class. These features are computed from the feature map generated by MobileNet, using a set of convolutional layers that are specific to each class.
The final step of the MobileNet SSD architecture is non-maximum suppression (NMS). NMS is used to eliminate redundant bounding boxes that overlap significantly with each other. The algorithm works by selecting the bounding box with the highest confidence score and removing all other boxes that overlap significantly with it. Repeat this procedure until there are no longer any overlapping boxes.

3.2 Description of Centroid Tracker

The Centroid Tracker algorithm works by maintaining a list of the most recent centroids for each tracked object and computing the Euclidean distance between the centroids of the new detections and the previous ones. Centroid Tracker is a simple object tracking algorithm that uses the centroid of objects to track their motion. It is particularly useful for tracking objects in real-time video streams. The algorithm works by associating centroids in subsequent frames of the video stream, using a distance-based matching algorithm.
In the context of object detection, the Centroid Tracker is typically used to track the motion of objects that have been detected by a detection algorithm such as MobileNet SSD. Once objects are detected in a video frame, the Centroid Tracker uses the coordinates of the bounding boxes around the objects to calculate their centroids. It then associates these centroids with objects detected in subsequent frames, based on their distance from the centroids in the previous frame. One of the advantages of the Centroid Tracker algorithm is that it is computationally efficient and can be implemented in real-time. It is also able to handle cases where objects move in and out of the frame, and where objects are partially occluded by other objects in the frame.
To use the Centroid Tracker, we first need to initialize it by creating an instance of the class and specifying the maximum number of frames to keep track of. Next, we iterate over each frame of the video stream, perform object detection on each frame using MobileNet SSD, and pass the resulting bounding box coordinates to the Centroid Tracker. The Centroid Tracker then uses these coordinates to calculate the centroids of the detected objects and track their motion over time. If the distance is below a certain threshold, the new detection is considered a match and associated with the corresponding tracked object. If the distance is above the threshold, a new object is created and tracked separately.

4. Results and Discussion

This section discusses the results given by MobileNet-SSD and Centroid-Tracking algorithms. The major input for this system is a video. This research project aimed to develop a people counter and location tracker to monitor crowds in real-time. The system was designed to be used in video surveillance settings, particularly in security and traffic monitoring applications, where people counting and tracking are crucial components. The system's ability to detect objects in real-time video and track them using background subtraction has made it a popular choice for real-time security

applications. The camera's range determines the area that can be monitored, and organizations such as banking institutions, jewelry stores, and the military can benefit greatly from this technology.
The process of detecting people in crowded scenarios involves feeding input images as shown in Fig 2 to our model. However, the input image needs to be preprocessed to meet the requirements of our model. The frame is resized to a maximum width of 500 pixels and changed from BGR to RGB to accomplish this.

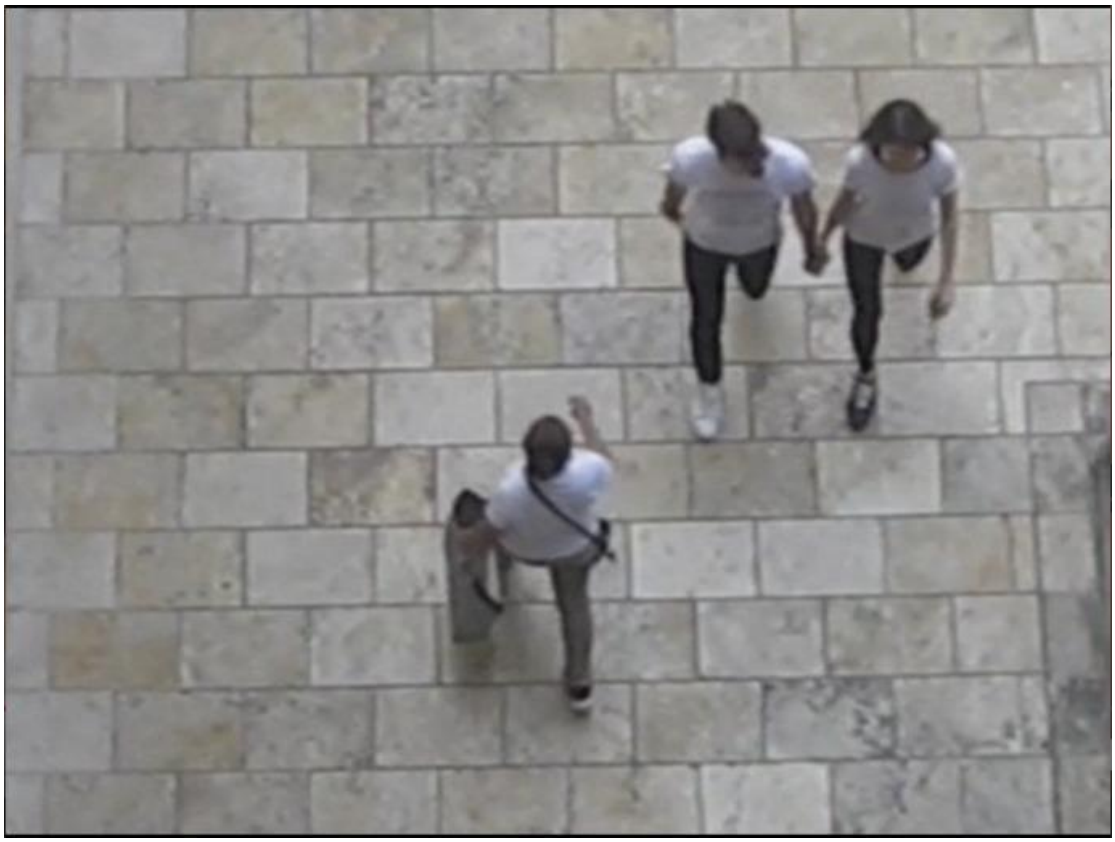

Figure 2: people moving in open space

The dnn module is then used to convert the preprocessed image to a blob. This blob is then passed to our model to generate predictions, which is a list containing seven floating values. At the first index of this list, we have the class ID of the detected object, and at the second index, we have the confidence or probability of the detection.

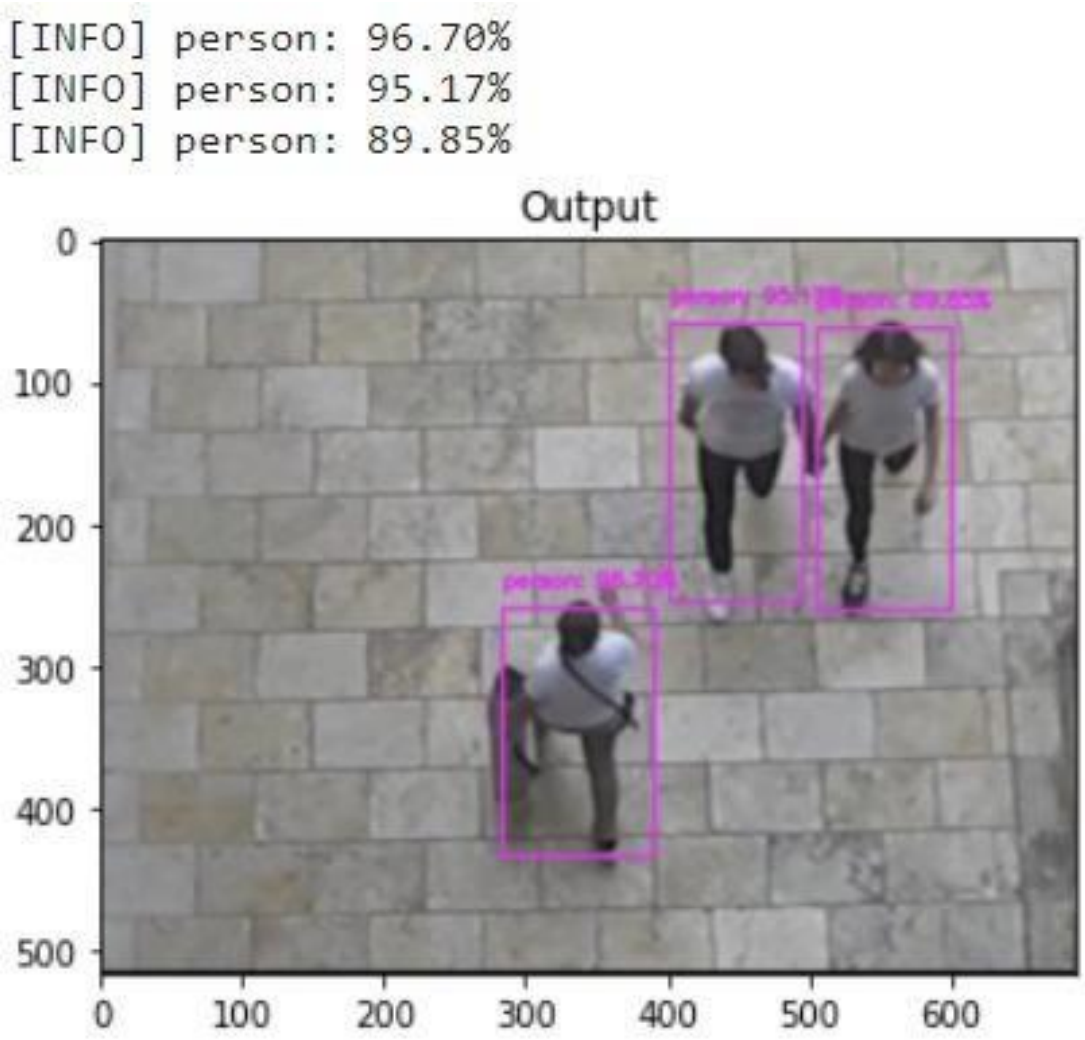

Figure 3: Detection of people in open space along with bounding box coordinates

It will detect the people in the images and give bounding box coordinates to each of the detected object along with the confidence value.

These bounding box coordinates are then passed as input to our centroid-based object tracker, which uses them to derive the centroids of the detected objects. By tracking these centroids, we can accurately count the number of people in a crowded scenario. Overall, this process involves image preprocessing, object detection using a deep neural network, and object tracking using centroid-based methods to accurately detect and count people in crowded scenarios. The CentroidTracker is a simple yet effective object tracking algorithm that uses the Euclidean distance between object centroids to track their motion across frames. The algorithm maintains a set of previously detected object centroids and matches them with the current set of detected centroids using the Hungarian algorithm, which finds the optimal assignment of object centroids based on minimizing the sum of their Euclidean distances.

The bounding box coordinates returned by the MobileNet SSD model are used to derive the object centroids, which are then tracked using the CentroidTracker module. The number of people in the room is determined by counting the number of unique object IDs that have been tracked over time as shown in the Fig.4 and Fig.5.
Once the object centroids have been matched, the CentroidTracker updates the position of the tracked objects and adds any newly detected objects to the set. This allows the algorithm to track objects across frames even if they are temporarily occluded or move out of view.

Figure 4: Tracking the movement of the person as he moves in and out

In the Fig.5, a prediction border has been given by the system, the centroid tracker has assigned an ID to the detected person as he crosses the prediction border. It updated the status to tracking and monitoring the count of total entered, exited and total people inside the room.

Figure 5: Tracking the movement of people

In fig.5, a person had been assigned the id1 as it assigned id0 to a person in fig.4, then it assigned id2 and id3 respectively to the people who were entering next. It incremented the enter count and total people inside to 2, as persons assigned id 0 and id1 crossed the prediction border. When the people assigned 1d2 and 1d3 crosses the prediction border, it increases the enter count and total people inside counter to 4.

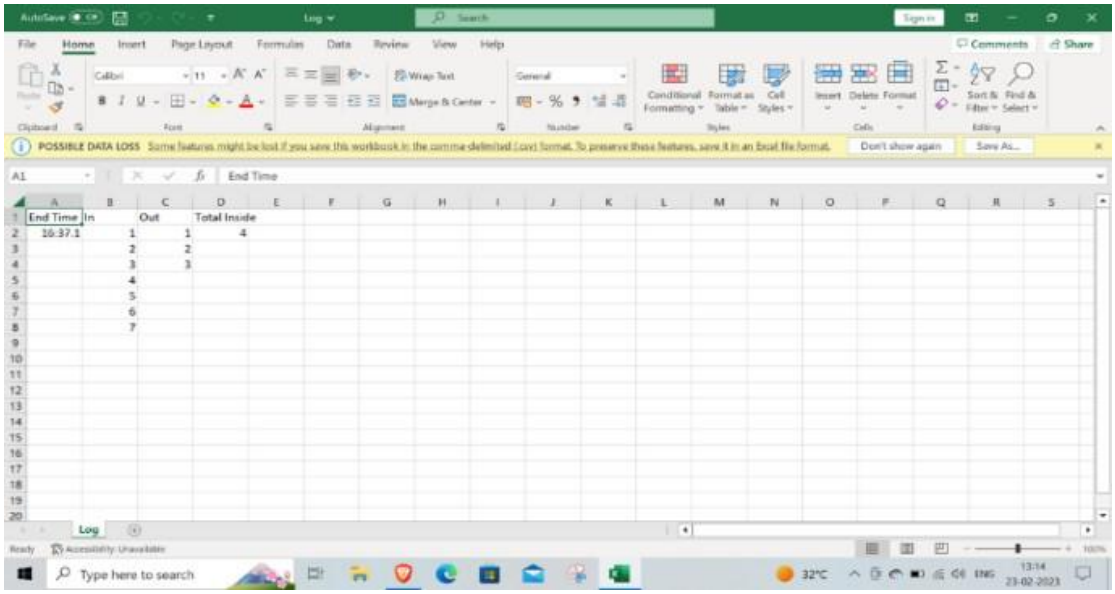

Figure 6: Record of count

The system will maintain a record of number of people entering and exiting the room as people enter and exit, the number of people present inside the room at a moment of time along with time-stamp is recorded as shown in Fig.6. The number of people entered the room and left the room and present in the room are recorded and maintained as a log of record along with the time.

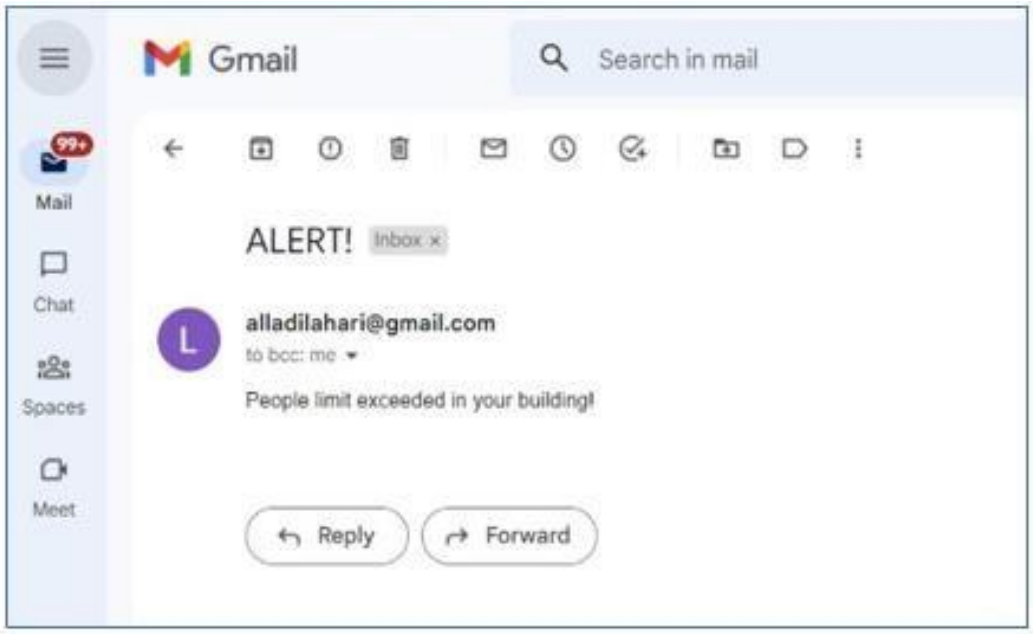

Figure 7: Sending alert mail to the user

A threshold value would be mentioned, if the number of people inside the room exceeds the mentioned threshold value, system will send an alert mail to the user saying people limit exceeded in your building using Simple Mail Transfer Protocol (SMTP). Threshold value can be given by the user based on the capacity of the room. This service helps the user to make sure there is no overload, without causing any inconvenience to the people inside the room.

5. Conclusions

In conclusion, people counting and tracking systems can provide valuable insights into how a business is operating, allowing for the development and implementation of strategies to optimize performance. These systems can count and monitor visitors as they enter and exit a physical space, providing separate reports for each. This data is especially useful for business leaders seeking to understand visitor behavior, including when they enter, how they move through the space, and when they exit. Additionally, people counting technology can show where visitors are spending their time and which products are attracting the most attention.

People counting systems in the retail sector can determine the precise number of visitors and monitor the proportion of those visitors who make purchases. This information is essential for understanding how effective marketing campaigns are at attracting visitors and converting them into customers. In shopping malls, people counting technology can help executives understand which areas of the mall are the most attractive and how traffic changes over time. This information can be used to shape marketing campaigns and plan ahead for peak periods. People counting systems are also useful in exhibitions and events, where they can help managers understand the peak hours of the event and optimize staff allocation accordingly. Additionally, people counting systems can provide total visitor numbers for events, which is essential for evaluating the success of the event and making plans for future events.

Object detection and recognition are critical components of people counting and tracking systems. These technologies are essential for real-time applications and can be used to detect any type of object in a physical space. By combining object detection and recognition with centroid-based object tracking, people counting and tracking systems can accurately track the movements of people in a video and count the number of people entering and leaving in realtime. This technology can be applied to a variety of scenarios, such as in retail stores, airports, and other public spaces where monitoring and tracking people's movements is necessary.

References

[1] D. Zhang, X. Du, and L. Zhang, “A Review on people counting and crowd density estimation”.

[2] S. L. Choi et al, “Multi camera people counting with deep learning-based object detection”

[3] K. Raja, S. S. S. Soumya,"A survey pf deep learning techniques for object detection and recognition".
[4] J. Guo, "pedestrian detection using recurrent neural networks".
[5] Joseph Redmon et al, "real time object detection with yolo".
[6] Afiq Harith Ahamad, Norliza Zaini, Mohd Fuad Abdul Latip, "Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI", 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2020
[7] D. Gozler, B. Isik and C. Topal, "A Real-time Queue Tracking Method for Waiting Time Estimation," *2022 30th Signal Processing and Communications Applications Conference (SIU)*, 2022
[8] J. Pegoraro and M. Rossi, "Real-Time People Tracking and Identification From Sparse mmWave Radar Point-Clouds," in *IEEE Access*, vol. 9, pp. 78504-78520, 2021
[9] Jia Wan, Qingzhong Wang, Antoni B. Chan, "Kernel-Based Density Map Generation for Dense Object Counting", IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 44, Issue: 3), 2022
[10] Kai Li, Chau Yuen, Salil S. Kanhere, Kun Hu, Wei Zhang, Fan Jiang, Xiang Liu, "An Experimental Study for Tracking Crowd in Smart Cities", IEEE Systems Journal (Volume: 13, Issue: 3), 2019 [11] M.I.H.Azhar, F. H. K. Zaman, N. M. Tahir and H. Hashim, "People Tracking System Using DeepSort," 2020 10th *IEEE* International Conference on Control System, Computing and Engineering (ICCSCE), 2020
[11] Shijie Sun, Naveed Akhtar, Huansheng Song, Chaoyang Zhang, Jianxin Li, Ajmal Mian, "Benchmark Data and Method for Real-Time People Counting in Cluttered Scenes Using Depth Sensors", IEEE Transactions on Intelligent Transportation Systems (Volume: 20, Issue: 10), 2019
[12] Weihong Ren, Xinchao Wang, Jiandong Tian, Yandong Tang, Antoni B. Chan, "Tracking-byCounting: Using Network Flows on Crowd Density Maps for Tracking Multiple Targets", IEEE Transactions on Image Processing (Volume: 30), 2020
[13] Xiaoheng Jiang, Li Zhang, Pei Lv, Yibo Guo, Ruijie Zhu, Yafei Li, Yanwei Pang; Xi Li, Bing Zhou, Mingliang Xu, "Learning Multi-Level Density Maps for Crowd Counting", IEEE Transactions on Neural Networks and Learning Systems (Volume: 31, Issue: 8), 2019
[14] Yuren Zhou, Billy Pik Lik Lau, Zann Koh, Chau Yuen, Benny Kai Kiat Ng, "Understanding Crowd Behaviors in a Social Event by Passive WiFi Sensing and Data Mining", IEEE Internet of Things Journal (Volume: 7, Issue: 5), 2020