Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Aabhas Bisaria
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
aabhasbisaria@gmail.com

Devesh Chauhan
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
deveshc2002@gmail.com

Harsh Varshney
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
harshjul11@gmail.com

Sushil Kumar
DSDL, Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
drsushil.cs@gmail.com

Umang Rastogi
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
rastogi225103@gmail.com

Vineet Sharma
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
vineet.sharma@kiet.edu

***Abstract*—The purpose of this study is to investigate the issue of increased electrical energy consumption in communal settings, with a specific focus on areas such as restrooms and multiplexes as the primary areas of investigation. Not only does the observed behavior increase the financial requirements of end-users and facility operators, but it also places a major strain on energy grids, which in turn leads to a rise in carbon emissions and an intensification of environmental concerns. A unique approach that makes use of the TensorFlow Object Counting API is proposed by the current inquiry as a means of addressing the concern that was highlighted earlier. The purpose of this strategy is to develop a system that can precisely count human subjects in real-time when they are present. To effectively decerase the amount of electricity that is not required, the primary objective of this cutting-edge system is to precisely identify and count the number of people that are present in shared spaces. When it comes to the development of advanced object counting systems, the open-source framework that has been developed on the TensorFlow is a platform that offers a method that is both user-friendly and efficient. Within the realms of item identification, tracking, and counting, the system exhibits capabilities that are both spectacular and resilient. By addressing the dual problems of rising electricity bills and negative effects on the environment, this technology offers a potential solution to the problem. Through this study, an attempt is made to investigate the factors that lead to excessive energy use and to propose a practical solution to this significant problem.**

***Index Terms*—Electrical Energy Saving, Human Detection, Single Shot Detector, Object Tracking Algorithm**

I. Introduction

In the current setting, one of the most important criteria that must be satisfied is to make the most of the available resources and to spend them as efficiently as possible. Currently, the problem at hand is the unregulated and ongoing consumption of electrical energy [1], which frequently surpasses the requirements that are being met. Because it raises a considerable concern, this constitutes a significant problem that calls for an immediate solution. There is a wide variety of various communal settings in which the observed predominance of wastefulness is apparent. Restrooms, public restrooms, and multiplexes are some examples of these types of locations; however, this list is not exhaustive. The analysis of the data reveals a discernible pattern in which the lighting fixtures located within the designated areas consistently tend to remain activated, even when there are no occupants present. This is particularly obvious in situations where there are no people around. Even when there are no other individuals in the vicinity, this is still the case.

To solve the issue of increased energy consumption, the primary objective of the current research is to offer a real-time human counter system as a means of tackling the problem [2]. The current system has been painstakingly constructed with the idea of accomplishing the primary objective of optimizing energy efficiency through the removal of superfluous electricity usage [3]. This target has been the fundamental inspiration for the construction of the system. When it comes to creativity and originality, the open-source framework that is currently being evaluated demonstrates an impressive amount of exceptional qualities. To construct this framework, the TensorFlow platform was utilized. The system provides users with a methodology that is not only user-friendly but also very efficient. This is done to simplify the creation of sophisticated object-counting systems. The system that is the focus of this investigation has an outstanding level of adaptability and efficiency in its operation. The resilient capabilities of the system in the areas of object recognition, tracking, and counting may be principally responsible for this particular consequence [4]. Taking into consideration the possibilities of this technology, there is a significant likelihood that it will bring about a revolutionary change in the management and operation of lighting systems in public spaces.

The technology that is being investigated demonstrates promising capabilities in properly recognizing the presence of individuals in public places and accurately monitoring the movements of those individuals. It is being taken into mind that certain skills are available. In addition to this, it will ensure that the lighting conditions within these specific regions are not only adequate but also excellent in terms of the amount of energy that they consume. When it comes to the study that we are conducting, our major purpose is not simply to cut down on the amount of electricity that is used [5]. The methodology that has been shown may be a unique approach that has the potential to result in considerable reductions in both the amount of energy that is used and the costs that are associated with both of these things.

A thorough analysis of our real-time human counting system is going to be carried out in the subsequent sections, which will thereafter be presented [6]. It will be important to perform a comprehensive examination of the technological components to properly execute the endeavor that has been recommended. This investigation will encompass the intricate mechanisms that are involved in the process of object recognition and tracking. As an additional point of interest, the primary objective of this research endeavor is to conduct a comprehensive examination into the combination of the system that was mentioned earlier with various techniques for the management of lighting. A thorough examination will also be carried out to evaluate the possible implications of this technology concerning the conservation of electricity in a range of public contexts [7]. This investigation will be carried out to examine the potential implications of this technology. The fundamental objective of the current study, which intends to provide in-depth information on these implications, is to provide a thorough understanding of the larger implications of our research in effectively addressing the existing challenges related to energy conservation. This is the primary objective of the study.

II. Related Works

The discipline of computer vision has recognized the significance of object detection, which is a key component that possesses tremendous utility across a wide variety of applications. There has been a noticeable shift toward the growth of real-time detection systems over the past ten years.The Single Shot Detector (SSD), first introduced in [8], represents a significant departure from traditional two-shot detectors such as R-CNN. The current study presents a new strategy that provides a methodology that can be used more effectively. By implementing a unified strategy that simultaneously combines regional proposal generation and object detection, the method known as SSD is able to achieve its goal. As a result of the strategy, an important paradigm shift took place in the respective field, especially in situations where convenience and efficiency are extremely important. One of the most innovative approaches to object recognition in computer vision was presented in the research paper ”You Only look Once: Unified, Real-Time Object Detection”. [6]. By adopting a unified model that analyzes the entire image at once, the work addressed the limitations of existing methods. The article was published in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. The success created a completely new paradigm for real-time object detection, marking a major departure from previous multi-pass methods. One of the main highlights is the You Only Look Once (YOLO) algorithm proposed in the paper. The image was first segmented into a grid, and then bounding boxes and class probabilities were predicted for each grid cell. This unified method eliminated the requirement for separate object location and classification, resulting in both faster and more efficient object detection. The simplicity and efficiency of the YOLO architecture was the basis for further research in the field.

One of the most innovative approaches to object identification in computer vision was provided in the research article titled “You Only Look Once: Unified, Real-time Object Detection” [6]. Through the introduction of a unified model that analyzed the full image in a single pass, the work addressed the constraints of existing methods. The article was published in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. As a result of the success, a completely new paradigm for real-time object detection was created, which marked a significant departure from the previously dominant multi-pass methodology. One of the most important highlights is the You Only Look Once (YOLO) algorithm that was proposed in the paper. The image was first segmented into a grid and then bounding boxes and class probabilities were predicted for each grid cell. This unified method eliminated the requirement for separate object location and classification, resulting in both faster and more efficient object detection. The simplicity and efficiency of the YOLO architecture was the basis for further research in the field.

A significant advancement in computer vision, particularly in instance segmentation, is demonstrated by the study ”Mask R-CNN” [9]. The Mask R-CNN model, based on the Faster R-CNN framework, presents a novel architecture that can predict object bounding boxes, class labels, and pixel-level masks simultaneously. This research paper examines the most important achievements, approaches and implications of the mask R-CNN model in the context of computer vision. The main contribution of the study is the ability of the Mask R-CNN paper to perform instance segmentation by extending the faster R-CNN architecture. In addition to the branches created for class classification and bounding box detection, this model also have a parallel branch for segmentation mask prediction. This technique made it possible to precisely define object boundaries at the pixel level, which was important for applications that require precise object placement and separation. In terms of integrating pixel-level mask prediction into an object detection framework, the innovation of Mask R-CNN marked a departure from traditional approaches that were used in the past. The challenges of accurately understanding spatial relationships through the simultaneous prediction of masks, jump boxes and class labels showed how versatile the model was. Subsequent research and applications in several fields have made Mask R-CNN a pixel-by-pixel segmentation industry standards.

One of the most important contributions to the field of object identification is the research article titled “Fast R-CNN” [10]. A unified framework that merged region proposal generation and object identification into a single model was established by Fast R-CNN to address the computational inefficiencies that were present in its predecessors. Girshick’s Fast R-CNN has had a notable impact on the field of object detection, inspiring following research and becoming a reference point for new algorithms. This research review examines the key principles, contributions, and impact of Girshick’s Fast R-CNN. As a result of the model’s potency in simultaneously handling region proposal and object detection tasks, more evolution have been inspired, such as Faster R-CNN and later versions. There are applications for Fast R-CNN in fields such as surveillance, automatic systems, and image-based searching application, which means that its influence goes beyond the realm of academics.

Feature Pyramid Networks for Object Detection address the challenge of differences in the field of object recognition [11]. After the authors realized how critical it is to detect features by capturing them at a variety of scales, they came up with the concept for the Feature Pyramid Network (FPN), which is an initialism for the framework. The most significant contributions, methodology, and impacts of FPN are investigated in this literature review, which focuses on object detection as its primary topic of investigation. The most significant addition that the study has made is the presentation of the design for the Feature Pyramid Network, which should be examined first and foremost. The formation of a feature pyramid, which is then employed to address the difficulty of scale variance, is accomplished by the utilization of a single convolutional network. Each of these feature maps gathers information at a specific scale, and their resolutions range from low to high. Features are the building blocks of this pyramid. The design, which enables high-level semantic properties to be paired with exact geographical information, makes it feasible to perform object recognition that is both more accurate and scalable. This functionality is made possible by the architecture.

III. Proposed System

In the system that is currently under investigation, two counting functions are considered to be of critical importance. The power to quantify the influx of individuals entering a certain location and the capability to monitor the count of individuals leaving that same area are both examples of these functionalities. To properly achieve the intended outcome, it is vitally necessary to focus one’s attention solely on the persons who have passed the threshold of intrusion. This is a requirement that cannot be avoided. By establishing the direction of the movement, specifically, whether it is oriented in a downward or upward trajectory, it is possible to make the completion of this aim easier to execute. By utilizing bounding boxes concept, which enable a more extensive visual representation, it is possible to generate a visual depiction of the detection of human. These bounding boxes [12] try to attempt enclose the human items that have been recognized as belonging to these categories. Additionally, a one-of-a-kind identification number is attached to each bounding box. This number serves the purpose of indicating the order in which the human artifacts were found. The architecture of the system has been painstakingly designed by the specifications through the use of a cascade method. The methodology that has been proposed begins with the identification of a human entity, which is then followed by the monitoring and tracking of that individual throughout a series of frames that are located close to one another. A high level of proficiency in the execution of counting precisely and tracking a large number of items at the same time is demonstrated by the methodology that has been proposed. Additionally, it can efficiently cope with situations in which a human-made thing that has never been seen before comes into view, and it immediately assigns it a distinct identity. This is a capability that it possesses.

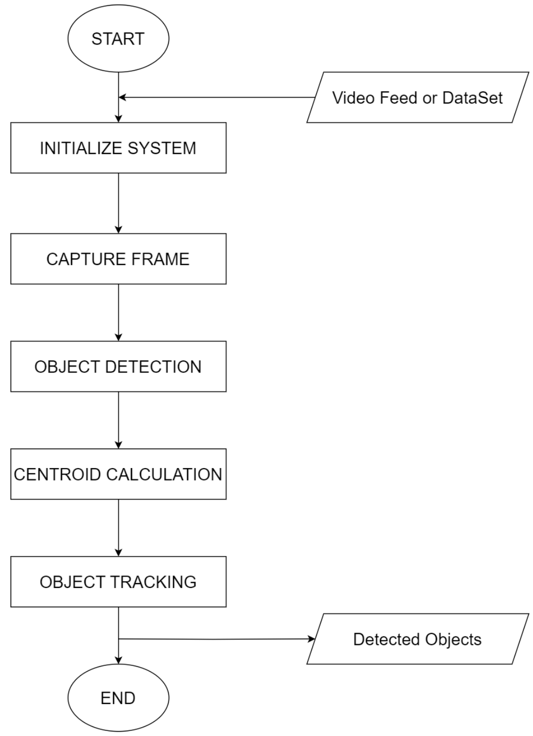

Fig. 1. Flow chart of proposed methodology

Therefore, to ensure the proper execution of this system, it is necessary to incorporate two essential components, which are as follows: both a human detection [13] module that is only devoted to the identification of human beings and a tracking mechanism that is exclusively designed to monitor the human objects that have been recognized are included in this system.

IV. Methodology

By utilizing a computer vision [14] investigation that is centered around the utilization of the Single Shot Detector (SSD) in conjunction with the Mobile Net architecture for real-time object detection [15], the researchers have used a computer vision investigation in the defining of the project. One of the most important feature of improving this func-

Algorithm 1 Object Tracking Algorithm

```
Initialize:
- List of detected objects
- Dictionary to store object IDs and their respective centroids
function DETECT_OBJECTS(frame)
    # Utilize Single Shot Detector (SSD) to detect objects
    detected_objects ← SSD.detect(frame)
    return detected_objects
end function
function TRACK_OBJECTS(detected_objects)
    for each object in detected_objects do
        if object is new then
            Assign a unique ID
            Add object ID and centroid to dictionary
        else
            Update centroid for existing object
        end if
    end for
end function
function DISPLAY_TRACKING_RESULTS(frame)
    for each object ID and centroid in dictionary do
        Draw bounding box around object using centroid
        Display object ID and bounding box on frame
    end for
end function
Main Loop:
while video is running do
    Capture frame from video
    detected_objects ← DETECT_OBJECTS(frame)
    TRACK_OBJECTS(detected_objects)
    DISPLAY_TRACKING_RESULTS(frame)
end while
```

Fig. 2. Pseudo code of object tracking algorithm

tionality of smartphone devices, IP cameras, and scanners is the establishment of this cutting-edge amalgamation, which efficiently optimizes both speed and efficiency demonstrated in Figure 3. In contrast to two-shot detectors like R-CNN, SSD can identify objects in a single iteration, which is a significant advantage over the latter. A sort of object detection [16] model known as the SSD is distinguished from two-shot detectors by the strategy that it takes. In contrast to two-shot detectors, which normally carry out the duties of region proposal and object identification in consecutive processes, SSD is capable of carrying out both functions concurrently. This indicates that SSD is capable of generating region proposals and detecting objects simultaneously over the network, which ultimately results in faster processing that is both accurate and precise. Real-time object detection [17] is made possible in a various of applications thanks to this approach, which removes the need for additional computing stages. As opposed to two shot detectors like R-CNN, the efficiency of SSD is remarkably higher. An important departure from the traditional multi-stage object detection methods is represented by the employment of SSD. By integrating region proposal and object detection, which is made possible by the utilization of SSD, the system increases both its efficiency and its speed. The ability to make decisions quickly, based on the recognition of objects inside a given scene, is required for real-time applications for them to function well. Centroid tracker is a well-established

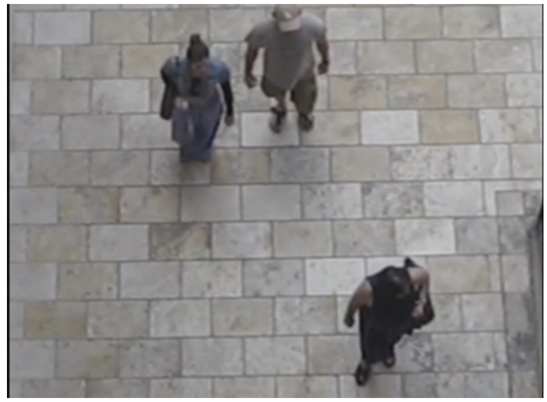

Fig. 3. Stock video frame without centroid allocation

method that is used to track objects across multiple frames. It is essential to make use of this tracking technique [18] to maintain regularity and difference over consecutive frames to guarantee a thorough evaluation of the items selected by SSD. The determination of the centroid of a bounding box is considered to be the core idea that is at play here. It is possible to notably increase the accuracy of object detection and tracking in visual images by establishing this computational strategy. A centroid tracker is used to determine the centroid of the bounding boxes that are produced by the Single Shot Multi-Box Detector (SSMD). To sum up and depict items that have been successfully detected inside specific frames of a given dataset or video sequence [19], bounding boxes inclusive of a set of two-dimensional coordinates, which are denoted as (x, y). Bounding boxes serve the aim of encapsulating and outlining things. To find the centroid of the bounding boxes, the tracker technique makes use of a computational approach. This makes it possible to discover the point within the frame that is the most central to the item. By making use of the concept of an object's centroid, it is possible to achieve the competence of monitoring and determining the trajectory and spatial coordinates of an object across number of video frames.

To ensure that the Centroid Tracker can perform its functions correctly and effectively, every object that is being observed must be assigned a unique and customized identity demonstrated in Figure 4. This enables the system to monitor and record the object's spatial displacement over consecutive frames. The tracking mechanism of the system is dependent on the unique identification (ID) [20] that is issued to each object. As a result of the system's consistent establishment of a correlation between the unique identification and the object in question, it can monitor and record the trajectory of the object throughout the procedure. In circumstances that involve several different items, the fact that it can independently track distinct things and provide a unique identity for each of them makes it extremely significant. The numerical IDs that have been provided are permanent, which ensures that continuity will be maintained and makes the process of reconstructing the line of travel of the object easier to accomplish.

Fig. 4. Stock video frame with centroid allocation through proposed methodology

V. Results Analysis

To accomplish the goals of this investigation, the system that is being studied is put through a series of tests on the graphics processing unit (GPU) of a computer that is located in the investigation area. The dataset that is being utilized in these experiments has been selected with considerable care. The central processing unit (CPU) was placed in the role of the target machine during the initial testing of the system. Regardless of this, it was quickly rejected because GPU displayed substantially higher levels of performance than previously thought. Only the number of line incursions going up and down are significant for counting. Tables I and II display the accurate manual count and the predicted count using the proposed technique for various camera angles. To get the final accuracy percentage, calculating the frequency of error by subtracting the anticipated count from the observed count. For the more accurate normal model, the ultimate accuracy is 91.66%, while for the faster model, it is 78.33%. Both results are elaborately outlined and shown separately below.

Rate of error = sum of frequency of error/(sum of anticipated increase and anticipated decrease) $\times$ 100%
= (0+1+1+1+2)/(1+4+6+3+11+11+2+6+2+4)$\times$ 100%
= (5/50) $\times$ 10 %
= 0.1000 $\times$ 100%
= 10%
Overall accuracy = total anticipated count - rate of error
= 100% - 10%
= 90%

Rate of error = sum of frequency of error/(sum of anticipated increase and anticipated decrease)$\times$ 100%
= (0+5+3+0+1)/(1+4+6+3+11+11+2+6+2+4)$\times$ 100%
= (9/50)$\times$100%
= 0.18$\times$100%
= 18%

Overall accuracy = total anticipated count - rate of error
= 100% - 18%
= 82%

TABLE I
Accuracy outcome of standard model

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	1	4
II	Anterior vision	6	3	7	3
III	Top vision	11	11	11	12
IV	Top vision	2	6	2	5
V	Top vision	2	4	1	3

TABLE II
Accuracy outcome with increased speed

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	2	3
II	Anterior vision	6	3	3	1
III	Top vision	11	11	10	9
IV	Top vision	2	6	2	6
V	Top vision	2	4	2	5

VI. Conclusion

The research embarked on a computer vision study aimed at leveraging the Single Shot Detector (SSD) approach alongside the Mobile Net architecture to enable real-time object detection. This integration significantly enhances speed and efficiency crucial for mobile devices, IP cameras, and scanners. Unlike two-shot detectors like R-CNN, SSD can identify objects in a single iteration, combining region proposal and object identification into a unified phase, eliminating the need for a separate region proposal network (RPN). This consolidation streamlines processing, optimizes resources and simplifies the detection pipeline. SSD's efficacy surpasses that of traditional two-shot detectors like R-CNN, offering a notable departure from multi-stage object detection techniques. Its benefits extend to various computer vision tasks, enhancing both efficiency and speed, crucial for real-time applications requiring swift decision-making based on object recognition within a scene.

This proposed system integrate a centroid tracker for object tracking across successive frames, ensuring consistency and clarity in evaluating objects identified by the Single Shot Detector (SSD). Centroid computation within bounding boxes enhances tracking accuracy and geographic localization of objects. By utilizing a centroid tracker alongside the SSD, the system computes centroids to encapsulate and track identifiable objects across frames. Each object is assigned a unique identification (ID) by the tracker, facilitating independent tracking and consistent recording of object movement over time. This unique identity system enables efficient tracking of various number of objects simultaneously by maintaining relationship between the unique identification number and the object, make it certify that it maintain uninterrupted record and simplifies the process for redeveloping object trajectories. This permanent nature of the identification number is streamlined for the tracking process, making it accurate easier to reconstruct object movements over time.

REFERENCES

[1] R. Roslina, A. Amelia, H. Pranoto, and B. V. Sundawa, "System of smart detection and control to electrical energy for saving of electrical energy consumption," 2021, pp. 2454–2465.

[2] H. Mokayed, T. Z. Quan, L. Alkhaled, and V. Sivakumar, "Real-time human detection and counting system using deep learning computer vision techniques," in *Artificial Intelligence and Applications*, vol. 1, no. 4, 2023, pp. 221–229.

[3] S. Chun and C.-S. Lee, "Applications of human motion tracking: Smart lighting control," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013.

[4] J. Liu, Y. Liu, Y. Cui, and Y. Q. Chen, "Real-time human detection and tracking in complex environments using single RGBD camera," in *2013 IEEE International Conference on Image Processing*, 2013, pp. 3088–3092.

[5] M. Elsisi, M.-Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. Darwish, "Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings," *Sensors*, vol. 21, no. 4, p. 1038, 2021.

[6] J. Redmon and et al., "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[7] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, and L. Soibelman, "Real-time activity recognition for energy efficiency in buildings," *Applied Energy*, vol. 211, pp. 146–160, 2018.

[8] W. Liu and et al., "SSD: Single shot multi-box detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I*, vol. 14. Springer International Publishing, 2016.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[10] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[11] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[12] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, "Bounding box regression with uncertainty for accurate object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2888–2897.

[13] W. Abd-Almageed, M. Hussein, M. Abdelkader, and L. Davis, "Real-time human detection and tracking from mobile vehicles," in *2007 IEEE Intelligent Transportation Systems Conference*, 2007, pp. 149–154.

[14] S. Lu, C. Hameen, and A. Aziz, "Dynamic HVAC operations with real-time vision-based occupant recognition system," in *2018 ASHRAE Winter Conference*, Chicago, 2018.

[15] O. Mata, J. I. Méndez, P. Ponce, T. Peffer, A. Meier, and A. Molina, "Energy savings in buildings based on image depth sensors for human activity recognition," 2023, p. 1078.

[16] A. P. Chandrasiri and D. Geekiyanage, "Real-time object detection system for building energy conservation: An IP camera based system," *THIRTY-FOURTH ANNUAL CONFERENCE*, 2018.

[17] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7991–8005, 2015.

[18] S. V. Viraktamath, M. Katti, A. Khatawkar, and P. Kulkarni, "Face detection and tracking using OpenCV," *The SIJ Transactions on Computer Networks and Communication Engineering (CNCE)*, vol. 1, no. 3, pp. 45–50, 2013.

[19] A. Pennisi, D. D. Bloisi, and L. Iocchi, "Online real-time crowd behavior detection in video sequences," *Computer Vision and Image Understanding*, vol. 144, pp. 166–176, 2016.

[20] C. Ning, H. Zhou, Y. Song, and J. Tang, "Inception single shot multibox detector for object detection," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017, pp. 549–554.