ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

An AI-Powered System for Detecting and Monitoring Crowds in Real-Time

[1]Mr. K.Lakshmi Narayana, [2]Gandi J V V Durga Ashok,

[1]Assistant Professor,Dept.of Master of Computer Applications, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2] Students,Dept.of Master of Computer Applications, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

***Abstract*—**

Surely, we can all take a page out of the COVID-19 pandemic's playbook and remember that each given infectious disease outbreak has the potential to trigger a global pandemic. The implementation of suitable crowd detection and monitoring technologies in public spaces is crucial for the prevention of epidemics and the improvement of healthcare delivery. One way to significantly reduce the incidence of new infections is to efficiently employ social distancing strategies. An concept for a social distancing system that uses crowd detection and monitoring in real-time was born out of this. For the purpose of better student monitoring within school grounds, this study suggests a completely autonomous system for Real-Time Crowd Detection and Monitoring. Created using OpenCV, this system can identify and count the number of individuals gathering at an instance using a Histogram of Oriented Gradients (HOG) and a Support Vector Machine (SVM) detector. If the number of persons in the cluster exceeds the permissible level, an alarm will be raised to remind everyone to follow the regulations. Some related terms are support vector machine (SVM), histogram of oriented gradients (HOG), object detection, crowd detection, and machine learning.

I. INTRODUCTION

Stations, malls, houses of worship, airports, public events, and many more locations often have large crowds. Because it provides the necessary data for crowd control, video surveillance is crucial in these areas. The detection and monitoring of crowds has many practical uses, such as ensuring public safety during pandemics, keeping an eye on traffic, limiting gatherings during emergencies, creating public spaces, studying crowd dynamics, preventing riots, and many more. Because of this, scientists have been working on models to help with things like counting, estimating densities, monitoring movements, and detecting behaviors. Identifying and categorizing various objects in a video stream as people is the first step in implementing this system for person detection. Using data and algorithms, a branch of AI known as machine learning teaches computers to behave like humans. A subfield of computer vision known as "Object Detection" is responsible for the categorization of various visual elements. Recognizing and classifying things (such as people, animals, trees, or cars) is what object detection is all about. In order to help computer-based vision algorithms understand "What objects are where," Object Detection delivers such information. This used to be done by humans, but it was very error-prone, computationally expensive, and took a long time. Modern computer algorithms for autonomous identification and tracking have reduced the need for human supervision, allowing for more cost-effective improvements in real-time performance. There are a lot of human detection algorithms out there. As seen in Figure 1, there are two categories of object detection techniques: those based on neural networks (deep learning) and those based on non-neural networks (conventional machine learning). Classical human identification techniques using non-neural methods include training models to differentiate between people and other objects in order to accomplish the pedestrian detection objective. These models are trained using artificially constructed feature extractors that extract important human features. Methods that do not rely on neurons include the Scale-invariant feature transform (SIFT) [1], the ViolaJones object identification framework employing Haar features [2], the Histogram of oriented gradients (HOG) features [3], and many more. By using many layers, such as an input layer, a hidden layer, and an output layer, the neural networks or deep learning method is able to recognize objects in images by identifying patterns. Both regression-based methods (e.g., YOLO and SSD) and region proposal-based algorithms (e.g., R-CNN, SPP-NET, and Fast R-CNN) are used in deep learning target identification approaches. Traditional object detection methods use less lines of code and are more successful than deep learning approaches when faced with a problem. Traditional algorithms are very general and apply the same logic to every picture. Deep neural network features, on the other hand, are trained on photos in the training dataset and, if not well crafted, will likely underperform on images outside of the training set. Training massive datasets can do these tasks, as shown, but it would be impractical and take too much time for a closed application. You may test your solution's efficacy outside of a training setting using traditional item identification approaches since they are fully transparent.

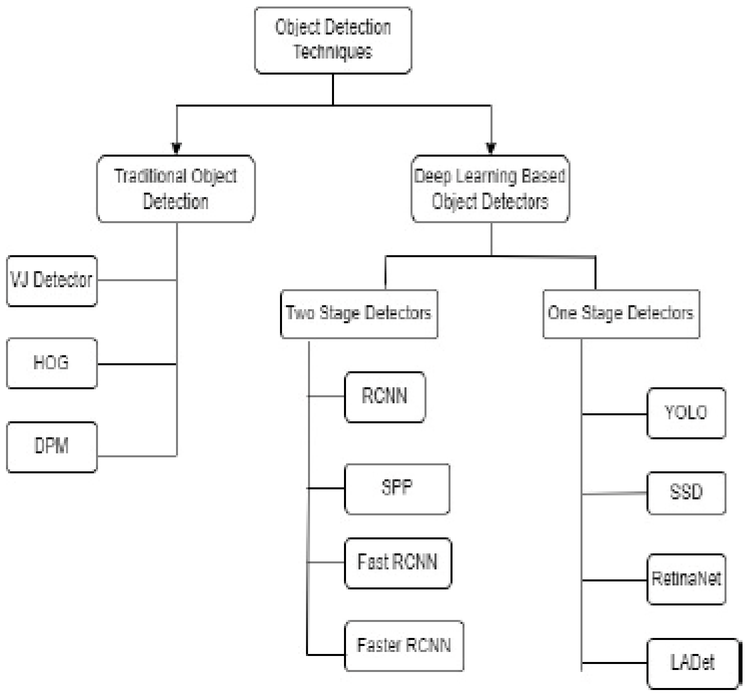

Fig. 1. Classification of Object Detection Methods

But there's a human detection algorithm in OpenCV. To identify people in photos and videos, it uses a Linear Support Vector Machine (SVM) and a pre-trained Histogram of Oriented Gradients (HOG) model. The basic human anatomy—two arms, two legs, a brain, etc.—is used to build AI models. After training, the model may be used to identify individuals in live video and still picture feeds. In order to determine how many individuals are present at any given time, this article uses an SVM classifier and HOG to analyze the results. There are five main parts to this paper: The previous work on crowd detection is covered in Section II, the approach for building a real-time CDMS is detailed in Section III, and the experimental findings are presented in Section IV. In Section V, we address final thoughts and potential future directions.

II. RELATED WORK

In order to accomplish pedestrian detection, several researchers have used various feature extraction and classification techniques. In order to aid in pedestrian identification, this research makes use of the work done in [4]. Based on studies, the HOG feature is the most discriminative standalone feature utilized in the literature [3]. The ability to precisely capture local edge/gradient information, together with its inherent resistance to changes in illumination, is a major selling point. Survey findings show that pedestrian identification using HOG has the best accuracy and the fewest false positives [5]. For optimal performance, almost all current detectors use HOG or a version of it in conjunction with SVM. Many classifications, particularly pedestrian classification, use Support Vector Machines (SVM) as a standard approach [6]. Many studies have attempted to estimate crowd size and count individuals, with published conclusions in the literature. Using HOG and SVM, the method suggested by [7] creates

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

a crowd detection system that sounds an alert in the event of aggressive crowd behavior. Using Light Weight Convolutional Neural Networks (LW-CNN), B. Vivekanandam's architecture [8] enables the implementation of crowd computing in any public venue, leading to enhanced counting accuracy. Since the COVID-19 pandemic began, researchers have used a variety of methods to study social distance. In order to control crowds and prevent the transmission of disease, algorithms for social distance categorization utilizing Deep Learning were developed [9] and [10]. Their system can identify when people are managing their distance from one another and will sound an alarm if they break the rules of social distance. The authors Yang et al. [11] suggested a consistent framework based on artificial intelligence monocular cameras to monitor social distance. The proposed method for controlling access to the target region and avoiding congestion makes use of a crucial social density. Thus, the literature study concludes that HOG is the best supervised classification standalone descriptor and that SVM provides the best performance with the least amount of work to implement. Chapter Three: Methods Fig.2 shows the suggested architecture for a campus-wide CDMS that monitors social distance in real-time. Video and photos are captured via IP cameras strategically placed throughout the campus. The server will save these video recordings in case someone wants to look at them later. The video scene's pedestrians are identified using machine learning algorithms. Background removal, feature extraction using HOG, and support vector machine (SVM) classification of identified pedestrians as human or non human are the main tasks that the system is expected to do. After humans have been detected, their counts are sent to the system that monitors social distance; if the norms are not followed, an alert will be raised.

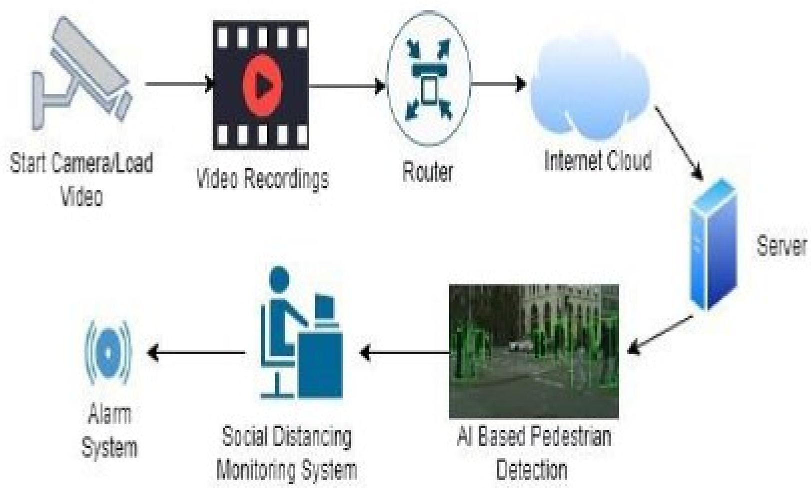

Fig. 2. Proposed Framework

A popular computer vision approach, Histogram of Oriented Gradients (HOG) [3][15][16] has a high success rate when it comes to object detection and feature extraction. To identify and characterize objects, HOG employs feature descriptors such as intensity gradients or edge directions. It uses a cellular division technique to create a gradient histogram for each pixel in a divided picture. The Sobel operator may be used to assess these gradients as seen in

$$Sx(y,x) = Y(y,x+1) - Y(y,x-1) \quad (1)$$

$$Sy(y,x) = Y(y+1,x) - Y(y-1,x) \quad (2)$$

equations (1) and (2):

the pixel intensity associated with the coordinates value (x, y) is denoted as Y(y, x). The horizontal gradient is represented by Sx (y, x), whereas the vertical gradient is shown by Sy (y, x). With the help of equations 3 and 4, we can get the gradient's magnitude (S) and direction ().

$$S = \sqrt{S_x^2 + S_y^2} \quad (3)$$

$$\theta = \arctan\left(\frac{S_y}{S_x}\right) \quad (4)$$

Cells are then classified based on the direction of the gradient. To determine the gradients along the horizontal and vertical axes, filtering is used. To apply filters horizontally, use [1, 0, 1], and to apply filters vertically, use [1, 0, 1]. As a description, we have the cumulative histogram of all the cells. To improve accuracy, normalization is applied to all areas in the detection window by providing a measure of the local histogram across larger windows of defined geographical regions. A feature vector is another name for this last vector, which has applications in object detection. The support vector machine classifier, which helps to identify whether the final product is human or not, is also built using these feature vectors.

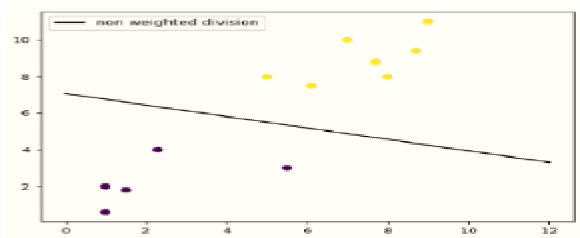

Fig. 3. : Linear SVM

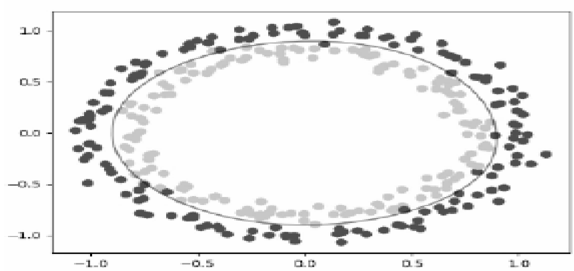

Fig. 4: Non-Linear SVM

In order to identify people clusters, a Real-Time CDMS monitors public areas. Figure 5 depicts the whole system process. The first step is to collect video sequences from the many security cameras that have been strategically placed. After that, we use Background subtraction to eliminate the picture's background components, leaving just the foreground objects, and then we extract image frames from the films. In order to guarantee social distance standards, the HOG and SVM algorithms detect persons and determine the number of individuals in a cluster.

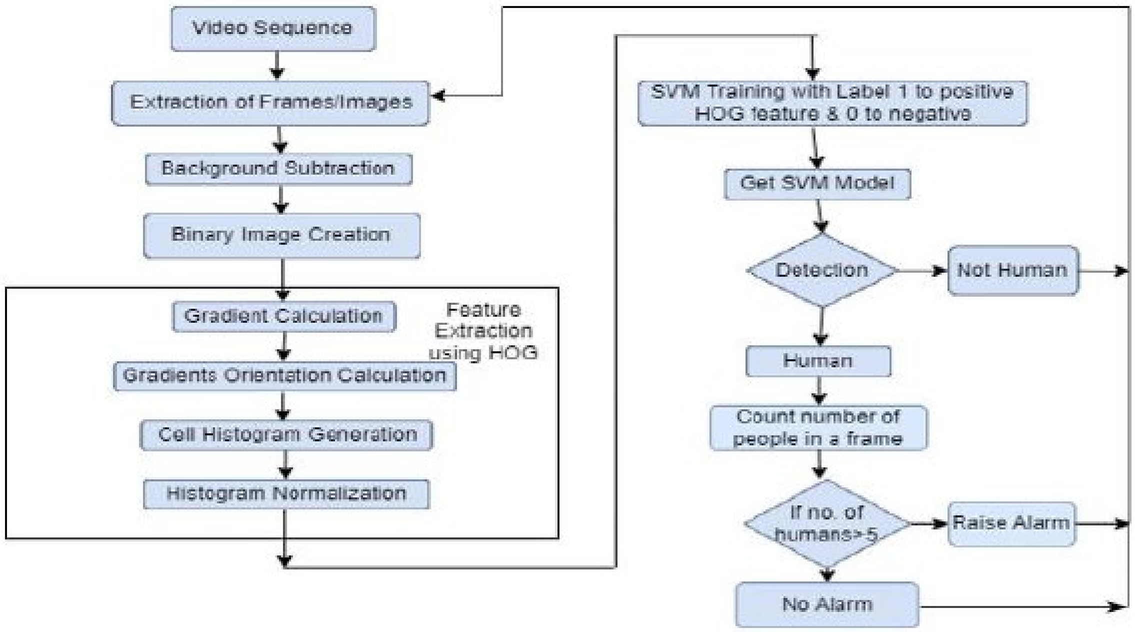

Fig. 5: Work Flow of Human Crowd Detection System

When the number of people in the cluster surpasses a certain level, the alarm goes off. Following the recommendations made in the COVID-19 guidelines, we have set the threshold value to 5 here [19, 20]. \

Algorithm to Detect Humans

Algorithm:

Step 1: cv2.HOGDescriptor() method is used and the descriptor object is set as HOGCV to extract the relevant information from the image.

Step 2: The SVM detector is initialized using the HOG Descriptor function

setSVMDetector(cv2.HOGDescriptor getDefaultPeopleDetector()).

Step 3: The function cv2.imread() is used to read images from a specified path.

Step 4: The function detectMultiScale(), is used to identify items in a picture and get their x, y, height, and width.

Step 5: We draw a bounding box around any recognized humans using the cv2.rectangle method.

Step 6: The count argument keeps track of the number of humans found.

IV. EXPERMIMENTS AND RESULTS

Specifically, the experiment trains and tests the model using the MIT pedestrian dataset [21]. Pedestrians in metropolitan environments are the subjects of 200 test shots and 509 training images, together with their left and right reflections. There are just a handful of positions and two camera angles (front or back). Accuracy is improved by tuning hyperparameters. The CMRIT college campus was the site of both the system's construction and the experiments. The method accurately predicted the human population under typical weather conditions. Figures 6 and 7 provide the screenshots of the output. Both the low-light and normal-light person detection scenarios are shown in Figures 6 and 7, respectively. The system ensured a secure environment to avoid diseases by checking the number of kids in a cluster and raising an alert if the count was discovered to be higher than 5.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Fig. 6: Detection in low light

Fig. 7: Detection in normal light

Table 1 displays the results obtained in various settings. In well-lit environments, the algorithm produced accurate student count predictions. However, the projected number of pupils was inaccurate when the light was dim. In wet circumstances, the system's student count predictions were spot on, but in foggy conditions, they were off. Therefore, in cases of poor illumination and adverse weather, there were occasional discrepancies between the expected and actual outputs.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

TABLE 1: OUTPUT PREDICTION UNDER DIFFERENT SCENARIOS

S.No	Output under different scenarios	Condition	Expected Output	Predicted Output
1	Image with 3 persons	Normal	3 persons	3 persons
2	Image with 3 persons	High Brightness	3 persons	3 persons
3	Image with 4 persons	Low Brightness	4 persons	3 persons
4	Image with 4 persons	Raining	4 persons	4 persons
5	Image with 4 persons	Fog	4 persons	3 persons

Table 2 shows the average test results as measures, which is a standard technique to illustrate the models' effectiveness. Accuracy is defined as the proportion of correctly predicted outcomes from the test data. The ability of a machine learning model to detect positive instances, or the amount of additional correct ones that were overlooked while the model displayed the correct ones, is called its sensitivity. The specificity of a model is defined as the proportion of false negatives it properly identifies. The ratio of correct forecasts to total positive predictions is called precision. To get an F1 score, take the harmonic mean of your recall and accuracy.

TABLE 2: PERFORMANCE METRICS

Measure	Values
Sensitivity	0.666
Specificity	0.918
Precision	0.40
Accuracy	0.90
F1 Score	0.498

Table 2 shows that the HOG + SVM combination achieves good accuracy and yields promising outcomes.

V. CONCLUSION AND FUTURE WORK

A crowd detection system was developed for an educational college using OpenCV and Python. The purpose of this model is to locate and tally individuals. We were able to do this by combining the SVM classifier that comes with OpenCV with the HOG Descriptor Algorithm, which allows us to identify individuals in photos. We measured a 90% success rate after testing the system in various difficult environments, such as low light, fog, rain, etc. A completely automated, real-time functioning surveillance system still has a ways to go before it can overcome some substantial challenges, even with the advancements in computer vision and related areas. From the more specific issues of camera location, installation, maintenance, and network bandwidth needs to the more general ones of camera longevity in typical weather and lighting conditions, installation costs, privacy concerns, and so forth, there is a wide variety of problems. In terms of accuracy, this research presents a good outcome for crowd detection. Using this method in conjunction with CCTV to screen people becomes possible during pandemics. Using this method as a foundation, mass screening may be implemented in densely populated areas such as train stations, transportation hubs, marketplaces, roadways, entrances to retail centers, schools, and colleges. Numerous exciting chances exist inside the system to broaden the scope of the endeavor and boost its efficacy. Counting people, identifying falls, monitoring public events, and spotting anomalous crowd behavior are just a few of the many other uses for live CCTV video. In order to improve the model's overall performance, future research may concentrate on developing a more efficient system that utilizes advanced Deep Learning algorithms to provide reliable results even in bad conditions.

REFERENCES

[1] Tony Linderberg. “Scale Invariant Feature Transform”, Scholarpedia, (2012)
[2] Viola P. and Jones M. “Rapid Object Detection using a Boosted Cascade of Simple Features “, CVPR (2001) Dalal N., Triggs B., and Schmid C. “Human Detection Using Oriented Histograms of Flow and Appearance”, Proceedings of European Conference on Computer Vision (2006)
[4] Yadav R. P, Senthamilarasu V, Kutty K, Vaidya V, Ugale S. P “A Review on Day-Time Pedestrian Detection” World Congress, SAE (2015)
[5] Solichin, Achmad & Harjoko, Agus & Putra, Agfianto. “A Survey of Pedestrian Detection in Video”, IJACSA (2014)
[6] Junil Tao, Klette R, “Vision Based Pedestrian Detection Improvement and Verification of feature Extraction Methods and SVM-Based Classification”, ITSC (2011)
[7] Singh, D. K., Paroothi, S., Rusia, M. K., & Ansari, M. A. “Human Crowd Detection for City Wide Surveillance”, Procedia Computer Science, (2020)
[8] Vivekanandam, B. "Speedy Image Crowd Counting by Light Weight Convolutional Neural Network." Journal of Innovative Image Processing (2021)
[9] Sangeeta Yadav, Preeti Gulia, Nasib Singh Gill, Jyotir Moy Chatterjee, "A Real-Time Crowd Monitoring and Management System for Social Distance Classification and Healthcare Using Deep Learning", Journal of Healthcare Engineering, (2022)
[10] Ansari MA, Singh DK. “Monitoring social distancing through human detection for preventing/reducing COVID spread”. Int J Inf Technol (2021)
[11] Yang, E. Yurtsever, V. Renganathan, K.A. Redmill, U. Ozg¨uner, “A vision-based social distancing and critical density detection system for covid-19”, arXiv e-prints pp. (2020)
[12] O. Javed, K. Shafique, and M. Shah.” A Hierarchical Approach to Robust Background Subtraction Using Color and Gradient Information.” Proc. IEEE Workshop on Motion and Video Computing, IEEE CS Press, (2002)
[13] Pushkar Protik Goswami, Dushyant Kumar Singh.” A hybrid approach for real-time object detection and tracking to cover background turbulence problem.” Indian Journal of Science and Technology (2016)
[14] Ruolin Zhang, Jian Ding, “Object Tracking and Detecting Based on Adaptive Background Subtraction”, International Workshop on Information and Electronics Engineering (2012)
[15] O. Kliper-Gross, T. Hassner, and Liu. Beyond.” Pixels: Exploring New Representations and Applications for Motion Analysis.” PhD thesis, Massachusetts Institute of Technology (2009)
[16] Surasak, Thattapon & Takahiro, Ito & Cheng, Cheng-hsuan & Wang, Chi-en & Sheng, Pao-you.” Histogram of Oriented Gradients for Human Detection in Video” ICBIR, (2018)
[17] Tzotsos, D. Argialas “Support Vector Machine Classification for Object-Based Image Analysis”, ISBN: 978-3-540-77057-2, (2008)
[18] Ilao, Joel & Cordel, Macario II. “Crowd Estimation using Region- Specific HOG with SVM”, JCSSE, (2018)
[19] https://economictimes.indiatimes.com/news/newsblogs/latest-dailynews- and-live-updates-december-24/liveblog/88464087.cms(Dec, 2021)
[20] https://www.outlookindia.com/website/story/india-news-coronavirusgathering- of-more-than-5-people-banned-will-lock-down-delhi-ifneeded- says-kejriwal/349213, (Mar, 2020)
[21] Papageorgiou and Poggio, “A trainable system for object detection”, IJCV (2000)