2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE2020), 21–22 August 2020, Penang, Malaysia

Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI

Afiq Harith Ahamad, Norliza Zaini, Mohd Fuad Abdul Latip
Faculty of Electrical Engineering
Universiti Teknologi MARA (UiTM)
Shah Alam Selangor, Malaysia
afiqharith05@gmail.com, drnorliza@uitm.edu.my, fuadlatip@uitm.edu.my

***Abstract*—In addressing the worldwide Covid-19 pandemic situation, the process of flattening the curve for coronavirus cases will be difficult if the citizens do not take action to prevent the spread of the virus. One of the most important practices in these outbreaks is to ensure a safe distance between people in public. This paper presents the detection of people with social distance monitoring as a precautionary measure in reducing physical contact between people. This study focuses on detecting people in areas of interest using the MobileNet Single Shot Multibox Detector (SSD) object tracking model and OpenCV library for image processing. The distance will be computed between the persons detected in the captured footage and then compared to a fixed pixels' values. The distance is measured between the central points and the overlapping boundary between persons in the segmented tracking area. With the detection of unsafe distances between people, alerts or warnings can be issued to keep the distance safe. In addition to social distance measure, another key feature of the system is detecting the presence of people in restricted areas, which can also be used to trigger warnings. Some analysis has been performed to test the effectiveness of the program for both purposes. From the results obtained, the distance tracking system achieved between 56.5% to 68% accuracy for testing performed on outdoor and challenging input videos, while 100% accuracy was achieved for the controlled environment on indoor testing. Whereas for the safety violation alert feature based on segmented ROI, it was found to have achieved better accuracy, i.e. between 95.8% to 100% for all tested input videos.**

Keywords—Covid-19, Person Detection; Social Distancing, Restricted Area, Segmented ROI

I. Introduction

Coronavirus is a large family virus that harms humans and animals. Covid-19 is known as a family member of coronavirus, first spread to Wuhan, China in December 2019. The outbreak then rapidly affected many countries in the world and had been declared as a pandemic by the World Health Organization (WHO) [1]. While many countries are still battling with Covid-19, the number of cases in Malaysia started to flatten [2]. Towards a flatter curve of Covid-19 cases in Malaysia, Malaysia Government has announced the Recovery Movement Control Order (RMCO) on 10th June 2020. Most of the economic sectors have been reopened and citizen of Malaysia are free to go out to do the daily routine with the terms of new normal. People that go outside must follow the guideline from the Ministry of Health Malaysia (MOH) and WHO to stop the spread of the viruses. One best practice known in stopping the spread of Covid-19 is by implementing social distancing between people with at least one meter away.

Based on the information from WHO, the coronavirus is spreading from a person to a person via small droplets from the nose and mouth. In other words, social distancing is the best practice where people can minimize physical contact with possible coronavirus carriers, by keeping the distance at least one meter away from each other. Figure 1-2 show the social measures taken in public places, with cross markings on the seats, that prevent people from sitting next to each other [3-4].

Fig. 1. Social distancing enforced on RapidKL trains.

Fig. 2. Social distancing enforced on Puteri Harbour International in Kota Iskandar Johor.

This study is proposed to support the actions on Covid-19 spread mitigation. It provides a solution for detecting people gathering in public places such as banks, shopping malls, clinics etc. The concept of person detection algorithm is used to accurately detect a person's presence in areas of interest and is then followed by measuring the distance between the detected persons.

In addition to social distancing measure, this study also includes detecting people in restricted or dangerous areas that will trigger a warning in the event of safety violation. Heavy transportation pathway, aircraft pathway, personal property, construction area and gas plant can be considered as important or hazardous regions commonly require visual surveillance [5]. Therefore, these areas need to be monitored to reduce the possibilities of people entry that will lead to unwanted incidents.

II. Research Background

Originally this study suggested the detection of persons in segmented regions especially for energy saving purposes, where lights in areas with no people can be automatically turned off. Therefore, some literature reviews have been done based on this initial purpose. However, the findings are still useful in reference to the new objectives since the basis of both studies is based on person detection in the region of interest (ROI).

A. System automation based on person detection

A study [6] presented automatic control of power supply in the classroom, where motion detection was used as the main indication for detecting human presence. Adopting image processing, background subtraction was used to detect the motion created by human presence. The method proposed is by differentiating between two images and frames in pixel level. Background subtraction process uses one image to be the reference image and act as a background image. The limitation of this method is when there is minimal motion such as strong wind and animals will also assume as "human presence".

[7] presented a study that uses RPi as the main component for controlling an automation system that is integrated with GSM for communication. This study is in the scope of home automation system that employs a centralized controller for controling the electrical appliances in the house. RPi has been a mainstream subject providing availability in dealing with automation. It can be used as a controller in any electronic and electrical automation with the wide flexibility in interfacing with many external electronic devices. With the customizable feature of RPi and the supported Python Programming language, it is suitable for automation as a variety of open-source libraries are available.

B. Location-based person detection

[8] focused on addressing the habits of users who always leave the classroom without switching off fans, lights and air-conditioner, which lead to the excessive usage of electrical energy. Arduino was used in this project as the controller, where it provides automatic lighting and control system for the classroom. The study proposed dividing the room into grids which control the lighting and fans in a particular area with human presence detection. Human presence detection used in this project is Passive Infrared Technology (PIR) as the input and placed it in a grid fashion and each grid represents different electrical appliances. The downside in using the PIR sensor is heat sensitive. It detects various sources of heat emitted including sun heat and any other radio-frequency radiation. Other than that, the emitted infrared radiation can be easily blocked by other object and it depends on the speed of the moving object which brings down the accuracy in detecting human presence [9]. This problem can be overcome by human presence detection in image processing to get higher accuracy in the detection of human presence.

[10] targeted university campus laboratories, offices and classroom to control the automation of lighting and air conditioning by detecting the presence and absence of humans. This study proposed an RSSI-based human presence detection system for energy-saving automation. Received Signal Strength Indicator (RSSI) is an indication for strength measurement of the transmitted signal by the access point. In this study, a human presence detection technique is developed by using the pattern of RSSI reading value. The system has a delay in the time response, where the module takes 40 seconds for 20 readings. The method proposed is suitable in small and moderate rooms as the WiFi signal propagation is better within the smaller space. Such limitation leads to the idea of using image processing as the human presence detection to gain more accuracy in detecting human.

[11] proposed to avoid overloading and elevator accidents while using the elevator. The research was conducted to count the real-time number of passengers in the elevator car. This study proposed a system with people counter in elevator car based on computer vision, image processing and human contour detection algorithm. Pattern noise elimination using image preprocessing technique was adopted to get better image quality and clarity. Background differencing is used to differentiate between the foreground and background which will detect the human body from the images taken. Based on the conducted analysis, the authors found the technique was unsatisfactory. This is because the contour of the human body may have contact shielding that makes the system unable to accurately count the person in the elevator car. Based on this study, we know that counting human and detecting human presence can be implemented using this technique and it is suitable especially in the large area such as lecture hall and laboratory.

[12] proposed to identify and localize the pedestrian from moving vehicle concerning road area and relative distance from the vehicle. They used the location classification to differentiate between the nonmoving background and moving foreground. The images classified into two regions which are the estimation road zone and the zone where should the pedestrian walk with the proposed walking human model. The dataset used in this project is extracted from Caltech and ETH [13], [14], internet and authors own images. The location of the image taken is divided into two regions which are the road zone obtained from the road lane boundaries. Location classification will determine and detect pedestrian from the system. From the research, the method authors used in their research can be utilized in detection human presence based on the region classification which can be implemented in dividing the lecture hall and laboratory area.

C. Pre-Trained Model

According to the performed studies, most existing object detection applications are using pre-trained CNN-based models. Therefore, this study aimed to use a similar approach, so further exploration has been done to determine the most suitable pre-trained model for this project.

In this context, [15] targeted to find the combination of speed and accuracy by comparing the performance of different object detection algorithms used in the convolutional neural network. The comparison is done with three different models which are Single Shot Detector (SSD), Faster Region-based Convolutional Neural Network (Faster R-CNN) and Region-based Fully Convolutional Neural Network (R-FCN). In a related study [16], the authors used Common Object in Context (COCO) dataset as the benchmark input. The experiment ran on Nvidia Titan X graphic card, 32 GB of Random-Access Memory (RAM) and power up with Intel Xeon E5-1650 v2 processor. The result of the experiment shows that Inception Resnet extractor model gives the max mean average precision (mAP) score of 30 with a min mAP score of 20. The frontier models test result will be shown in Table I. As far as we can see, the faster model is SSD MobileNet with test dev mAP score of 19 and the most accurate is Faster R-CNN Inception Resnet with score 34.2. Slightly below the score, Faster R-CNN Resnet-101 Model with good balance result. This project gives a good overview in comparing pre-trained model from the various algorithm in term of time rate and accuracy. It is useful in determining which pre-trained model is suitable to be implemented for the proposed idea.

TABLE I. MAP SCORE RESULT OF FRONTIER MODELS ON COCO DATASET

Model	Mini validation mAP	Test dev mAP
SSD MobileNet	19.3	19
Faster R-CNN Resnet-101	31	30.9
Faster R-CNN Resnet-101	33.2	33
Faster R-CNN Inception Resnet	34.7	34.2
R-FCN MobileNet	13.8	13.4

For model selection, a comparison has been made on several pre-trained models with various data sets such as COCO, Kitti and Open Images. The comparison was done to evaluate the time taken for execution, accuracy and number of objects detected with selected Test Images Set. A paper [15] discussed a comparative analysis of variegated pre-trained models for discrete class-labels. COCO object detection model has been used in comparing the images data set. Result from this finding, Mask R-CNN Inception Resnet version 2 Atrous and Faster R-CNN Inception version 2 COCO get the highest accuracy with 99% accurate over the 27 test images. The results of this comparison are shown in Table II. Between the two models, Mask R-CNN Inception Resnet version 2 Atrous always get the higher images detected. Considering the accuracy and the time taken for the execution, SSD MobileNet V1 COCO is better for the proposed idea in detecting object as we want the system to be accurate as possible with a good time rate for the execution.

TABLE II. COMPARISON RESULT BETWEEN FOUR OBJECT DETECTION MODELS

Model Name	Execution Time (s)	Highest Accuracy	Object Detected
SSD Mobilenet V1 COCO	219.58	94%	2
SSD Inception V2 COCO	298.22	97%	2
Faster RCNN Inception V2 COCO	420.71	99%	3
Mask RCNN Inception Resenet V2 Atrous	6008.02	99%	5

According to the findings, the data obtained shows that the MobileNet SSD object detection model has a faster execution time for object detection. Therefore, due to hardware limitation, MobileNet SSD object detection model will be used in this study. However, it is known that the percentage of accuracy will be compromised in this study.

III. METHODOLOGY

In this study, the proposed idea is developed based on Python 3, OpenCV and Caffe framework. OpenCV library is used to utilize the image processing methods that will be explained further in this section.

The main purpose of this system is to process captured video footage for person detection and further processing for social distancing or safety violation. So, the process starts with reading the frames of a video feed one by one. This is shown in Fig. 3 which illustrates the whole sequence of activities in a flowchart. The most important feature of this study is the object detection framework. This is due to the element of this study that focuses on determining the location of a person from the input frame. Hence, choosing the most suitable object detection model is important to avoid any problems in detecting persons.

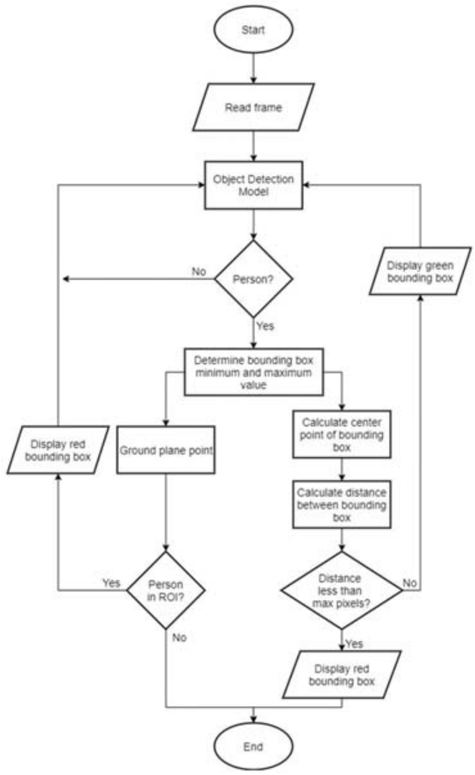

Fig. 3. Person detection for social distancing and safety violation alert based on segmented ROI flowchart.

A. Object Detection Model

In this study, Caffe deep learning model framework is used to run the object detection model. The model chosen is MobileNet SSD due to the short time taken for the execution.

B. Threading Parallelism

In Python 3, threading allows the different part of the program to run concurrently. In this study, using threading will improve the execution time to process the object detection on each frame. Multithreading approach will be used to run the frame and processing the object detection at the same time.

C. Masking frame for ROI area estimation

Masking is a technique in image processing which define as a small image piece and use it to modify a larger image. Masking involves setting some of the pixel values in an image to zero and some other background value as in (1).

$$f_{mask} = f - \text{mask}(f) \qquad (1)$$

It will isolate the image, *f*, by masking the area of ROI[17]. Video for an instance is a series sequence of images that been play in a certain amount of time. In this study, OpenCV masking method will be used to create ROI for each frame of the input frame.

D. Determine person location

In determining the position of a person's bounding box as well as the segment involved, each ground plane point is used to compare the ROI range.

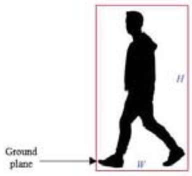

Fig. 4. The ground plane of a human walking model.

Surveillance cameras are usually placed at high places as the overhead camera especially to monitor a certain area e.g. high-risk area or areas of interest for an organization. In this case, it is more suitable to compare the ground plane for the detection box instead of using the center point value.

E. Calculate the center point of a bounding box

To measure the center point, $C(x, y)$, of the bounding box for the detected person, midpoint equation is used as in (2).

$$C(x, y) = \left(\frac{x_{min}+x_{max}}{2}, \frac{y_{min}+y_{max}}{2}\right) \quad (2)$$

Each of the minimum and maximum value for the corresponding width, x_{min} and x_{max}, and height, y_{min} and y_{max}, of the bounding box will be used to calculate the center point of the bounding box.

F. Calculate distance between bounding box

To measure the distance, $C_1(x_{min}, y_{min})$ and $C_2(x_{max}, y_{max})$, between each of the detected person in the frame, distance equation is used as in (3).

$$d(C_1, C_2) = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2} \quad (3)$$

In this study, the center point of the bounding boxes is taken to determine between two different locations of the bounding boxes. After getting the center points value, the algorithm will calculate if the distance is lower or higher than 300 pixels.

IV. Results and Discussions

System development for this project has been completed based on Python 3, OpenCV for image processing techniques and Caffe object detection model framework. Based on this developed system, some analysis has been performed to test its effectiveness and results have been obtained. Specifically, the MobileNet SSD Caffe model has been used in this study as the key algorithm in person detection. For program tweaking, the main video footage is captured from the entire scene set in the living room, where the camera is positioned high to gain overhead view.

A. Masked ROI

As a step in determining areas of interest, the masked ROI is made to consists of two parts. As shown in Fig. 5., the first masked area is the foreground of the frame. The foreground indicates the ROI for the area where the person detection will trigger a return of the desired ground plane point value. Then, the foreground mask is inverted to get the background mask. Fig. 6. shows both foreground and background frames are combined into a single frame to get the desired ROI.

Fig. 5. OpenCV masking step to get foreground frame.

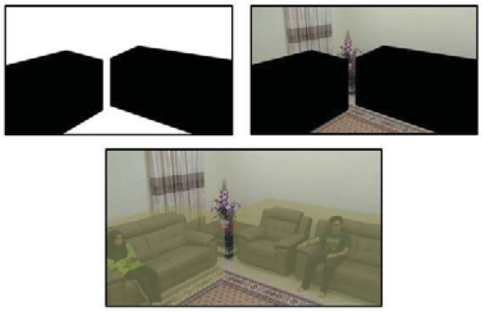

Fig. 6. Masked background frame with combination of foreground frame.

B. Person Detection Localization and Social Distancing

MobileNet SSD Caffe model consists of 20 object classes. The targeted object is "person", which is the 15th class of all object classes available. In this study, the person detection algorithm consists of two parts. The first part is configured to compare the distance between the computed center point of the bounding boxes for each detected person. The distance will be compared with the default social distance range. The default pixels for social distance varies on the frame input.

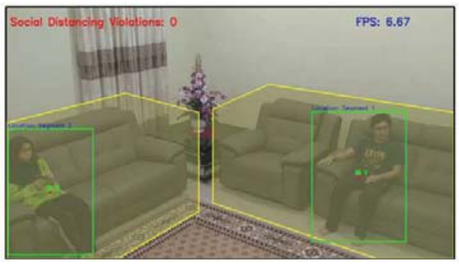

Fig. 7. Green bounding boxes when there is no social distance violation.

Fig. 7. shows the green bounding boxes for each possible detected person when the distance for each possible detected person bounding boxes is longer than the default value of social distance range. The experimental default pixels value shown in Fig. 7. is 300 pixels. The default of the minimum distance for social distance varies on the different video input as the camera perspective view is different. In contrast, when the safe social distance is violated, the bounding boxes of affected persons are

changed to red indicating that the distance is below the minimum value for a safe distance. In this way, the difference between those who violate the safe distance and those who do not can be seen. The results are shown in Figure 8.

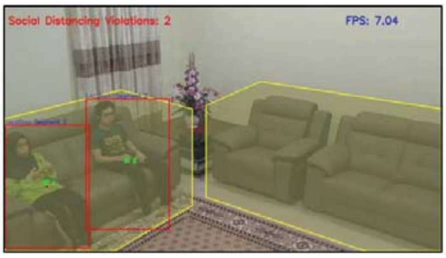

Fig. 8. Red bounding boxes indicate an alert or warning when there are social distance violations.

To measure the effectiveness of this system, the accuracy is calculated. In calculating the accuracy, the values for true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for social distance monitoring are counted. The formula used in calculating the accuracy is shown in (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The social distancing monitoring system accuracy is tested on four different videos, that include dataset from PETS2009, TownCentre, VIRAT_S and self-taken video[18]–[20]. Self-taken video is recorded in an indoor and controlled environment. Hence, the highest accuracy for the system is based on a self-taken video. Table III shows the computed accuracy for this system.

TABLE III. ACCURACY OF SOCIAL DISTANCE MONITORING.

No	Video	TP	TN	FP	FN	Accuracy
1	Self-taken	2	2	0	0	100%
2	TownCentre	11	19	14	4	62.5%
3	PETS2009	14	38	19	5	68%
4	VIRAT_S	9	4	0	10	56.5%

Based on the findings, this study could not maintain the highest accuracy in the social distancing system. This problem is caused by the object detection model that cannot detect the presence of persons in some of the video. In a difficult environment, the object detection model can hardly detect people's presence and accurately guess the exact location of a person. This was found to have occurred in several frames in the video involved. Based on the observation made, the object detection model can detect the presence of a person if the camera used to capture the video is placed close to the object or in a controlled indoor environment and situation as shown in Fig. 7 and Fig. 8. Therefore, the social distancing system needs to be improved for the outdoor environment especially for videos that capture distant scenes.

Fig. 9. and Fig. 10. shows the object detection framework fails to locate the person accurately in the frame. This problem is causing difficulties and inaccurately measure the distance between each bounding box because of the social distance measurement is depending on the detection box center point values.

Fig. 9. Multiple red bounding boxes for a single person show the object detection fails to locate person accurately on dataset PETS2009.

Fig. 10. The object detection framework fails to locate the location of the person in the frame on dataset TownCentre.

Whereas, the second part of person detection algorithm in this study involves localization of detected people in entering the prohibited or dangerous area. The object detection algorithm is set up to get the coordinates of the person bounding boxes which are used to determine the ground plane of the bounding boxes for the detected person. The ground plane value for the bounding boxes will be compared with the four points of the ROI to detect if a bounding box is inside the ROI. Fig. 11. shows when the bounding boxes are inside the ROI, which indicate that the persons are in the prohibited or dangerous area. The output as shown in Fig. 7. and Fig. 8. also show the same approach; where there is a label above each bounding box indicating the segment in which it is located. In contrast, when a detected person is outside the ROI, the system will not display and return the red bounding boxes for the output, which indicates no violation has occured or no people are detected within the restricted or dangerous area. Fig. 12. shows the output frame for the condition.

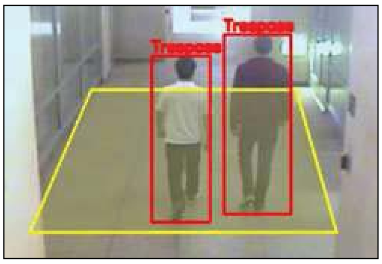

Fig. 11. Red bounding boxes with the word "Trespass" indicate an alert or warning when entering the ROI.

The person detection in the area of interest or dangerous is tested on CamNeT dataset and self-taken video [21]. The accuracy of the system is shown in Table IV. The outputs themselves can be seen in Fig. 7 and Fig. 8 that show the location of the detected person at the top of the bounding box. Based on the computed accuracy, this study has achieved high accuracy for the safety violation alert based on ROI system. This system uses the same object detection algorithm as the social distancing monitoring system, but it managed to produce more accurate results. From the observations made, both video inputs used to test this feature have used internal visual input and the location of the camera is close to the monitored ROI. As such, it can accurately locate the person's location.

Fig. 12. No red bounding boxes on detected persons outside the ROI.

TABLE IV. ACCURACY OF ALERT FOR RESTRICTED AREA BASED ON ROI.

No	Video	TP	TN	FP	FN	Accuracy
1	Self-taken	2	2	0	0	100%
2	CamNeT	55	58	0	5	95.8%

V. CONCLUSION

Social distancing is one of the important precautions in reducing physical contact that may lead to the spread of coronavirus. Consequences of non-compliance with these guidelines will be causing the higher rates of virus transmission. A system has been developed using Python and OpenCV library to implement two proposed features. The first feature is on detecting violations of social distancing, while the second feature is on detecting violations of entering restricted areas. Both features have been tested for accuracy. Based on the overall results, this study is seen to meet all of its objectives. However, there are some limitations to the results obtained. Based on the tests performed on the system, the results show that the object detection model used for detecting persons is having the difficulty in detecting people correctly in the outdoor environment and difficult scenes with distant scenes. For further improvement in the future, a better object detection model can be implemented.

ACKNOWLEDGEMENT

We would like to extend our acknowledgement to Universiti Teknologi MARA (UiTM) and to those who have directly and indirectly contributed to our project. This research is funded by the Ministry of Higher Education Malaysia via the Fundamental Research Grants Scheme Research (FRGS/1/2017/TK04/UITM/02/26).

REFERENCES

[1] World Health Organization, "Coronavirus Disease 2019," *Coronavirus disease (COVID-19) pandemic*, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed Jun. 19, 2020).

[2] M. Ministry of Health, "COVID-19 MALAYSIA," *Http://Covid-19.Moh.Gov.My/*, 2020. http://covid-19.moh.gov.my/.

[3] H. H. Naz, "Social distancing enforced on all RapidKL trains," *NST*, 2020. https://www.nst.com.my/news/nation/2020/03/577603/social-distancing-enforced-all-rapidkl-trains (accessed Jun. 30, 2020).

[4] The Star, "'Be ready for stricter MCO,'" *The Star*, 2020. https://www.thestar.com.my/news/nation/2020/03/27/be-ready-for-stricter-mco (accessed Jun. 30, 2020).

[5] Australian Government, "Prohibited, restricted and danger areas _ Civil Aviation Safety Authority." https://vfrg.casa.gov.au/operations/cruising-level-requirements/prohibited-restricted-and-danger-areas/ (accessed Jun. 27, 2020).

[6] D. Ganiger, K. A. Patil, P. Patil, and M. Anandhalli, "Automatic Control of Power Supply in Classroom Using Image Processing," in *Proceedings - 2017 International Conference on Recent Advances in Electronics and Communication Technology, ICRAECT 2017*, 2017, pp. 230–234, doi: 10.1109/ICRAECT.2017.16.

[7] S. Khedkar and G. M. Malwatkar, "Using Raspberry Pi and GSM Survey on Home Automation," in *International Conference on Electrical, Electronics, and Optimization Technique, ICEEOT 2016*, 2016, pp. 758–761.

[8] S. Suresh, H. N. S. Anusha, T. Rajath, P. Soundarya, and S. V. P. Vudatha, "Automatic lighting and Control System for Classroom," in *Proceedings of 2016 International Conference on ICT in Business, Industry, and Government, ICTBIG 2016*, 2017, doi: 10.1109/ICTBIG.2016.7892666.

[9] V. S. Rao and S. S. Kowshik, "PIR Sensor Characterization and a Novel Localization Technique using PIRs," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 2015.

[10] M. Habaebi, R. Rosli, and M. R. Islam, "RSSI-based human presence detection system for energy saving automation," *Indones. J. Electr. Eng. Informatics*, vol. 5, no. 4, pp. 339–350, 2017, doi: 10.11591/ijeei.v5i4.356.

[11] H. Fan, H. Zhu, and D. Yuan, "People Counting in Elevator Car Based on Computer Vision," in *IOP Conference Series: Earth and Environmental Science*, 2019, vol. 252, p. 032131, doi: 10.1088/1755-1315/252/3/032131.

[12] J. Hariyono and K. H. Jo, "Detection of pedestrian crossing road," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2015-Decem, no. January, pp. 4585–4588, 2015, doi: 10.1109/ICIP.2015.7351675.

[13] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, 2012, doi: 10.1109/TPAMI.2011.155.

[14] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, vol. 2009 IEEE, pp. 304–311, 2009, doi: 10.1109/CVPRW.2009.5206631.

[15] N. Yadav and U. Binay, "Comparative Study of Object Detection Algorithms," *Int. Res. J. Eng. Technol.*, pp. 586–591, 2017, [Online]. Available: www.irjet.net.

[16] T. Y. Lin *et al.*, "Microsoft COCO: Common objects in context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.

[17] W. K. Pratt, *Digital Image Processing*, vol. 19, no. 3. 1994.

[18] S. Oh *et al.*, "AVSS 2011 demo session: A large-scale benchmark dataset for event recognition in surveillance video," *2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011*, no. 2, pp. 527–528, 2011, doi: 10.1109/AVSS.2011.6027400.

[19] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," *Proc. 12th IEEE Int. Work. Perform. Eval. Track. Surveillance, PETS-Winter 2009*, pp. 0–5, 2009, doi: 10.1109/PETS-WINTER.2009.5399556.

[20] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3457–3464, 2011, doi: 10.1109/CVPR.2011.5995667.

[21] S. Zhang, E. Staudt, T. Faltemier, and A. K. Roy-Chowdhury, "A camera network tracking (CamNeT) dataset and performance baseline," *Proc. - 2015 IEEE Winter Conf. Appl. Comput. Vision, WACV 2015*, pp. 365–372, 2015, doi: 10.1109/WACV.2015.55.