Social Distancing Monitoring for Real-Time Deep Learning Framework

Sunil S. Harakannanavar, R. Pramodhini, A. S. Sudarshan, Atish, Rohit Jadhav, and S. Kruthik

1 Introduction

An effective measure against the novel COVID-19 pandemic is social distancing. However, in a real sense for the public, it is not possible for them to estimate the distance mentally. An automated surveillance system can aid the augmentation of the individual's perspective capability. Employing a dynamic monitoring system brings in a lot of ethical concerns. The privacy of the individuals is at stake if data is recorded and stored. The monitoring system is expected to be instantaneous without any capacity for data storage. Additionally, the detector must not non-favor any object class during identification. This can be attained by developing a system based on deep learning. A crucial aspect over here is that the system is supposed to be non-intrusive meaning that the warning system cannot target the individuals directly. This can be done by displaying visual cue if the subjects in the frame violate COVID-19 norms. This is a vital for establishment between surveillance and society. After finding the gaps in the literature survey, the existing systems are found to be inefficient when it comes to real-time implementation. The deep learning model for detecting humans in the frame and creating bounding boxes around them to determine the distance is proposed. Further, detections are represented using spatial coordinates. If the measured distance between two or three people is less than the assigned value, the system will display a visual cue. In the previous outbreak, a lot of families lost their loved ones and until vaccines prove to be completely efficient, maintaining social distance is an unswerving solution to the pandemic. Any form of social gathering like weddings, parties, or family events was restricted for the public and following the COVID-19 protocol everywhere was not feasible for people, especially for senior citizens who were more susceptible to the infection. People went outside carrying

S. S. Harakannanavar (✉) · R. Pramodhini · A. S. Sudarshan · Atish · R. Jadhav · S. Kruthik
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore, Karnataka 560064, India
e-mail: sunilsh143@gmail.com

N. R. Shetty et al. (eds.), *Advances in Communication and Applications*, Lecture Notes in Electrical Engineering 1105, https://doi.org/10.1007/978-981-99-7633-1_5

a fear inside them of the invisible virus which took a toll on their mental health. According to the WHO, maintaining social distance is proven to be 98% efficient and they insist on following the norms even after pandemic. Hence, what people lacked was a surveillance system that would detect whether people are following the measures or not and thereby giving them a sense of security.

The paper is organized as follows. Section 2 describes the descriptions of exist sting model. The proposed model is discussed in Sect. 3. The result analysis of the proposed model is done in Sect. 4. The conclusion and future research are addressed in Sect. 5.

2 Related Work

In this section, the work related to existing methodologies used for Social Distancing Monitoring Framework Machine learning algorithms are discussed. Yadav et al. (2022) developed a model to monitor and manage the crowd for social distance. The YOLO and deep sort techniques are used to evaluate the model. Ahmed et al. (2021) discussed a platform related to deep learning, the YOLO was used for object recognition. Mangshor et al. (2021) introduced the YOLO algorithm which was implemented with the SDLC to recognize an object. The deep learning model was developed by Gopal and Ganesan (2022) to measure the distance between the crowds. Here, ED is used for classification process. The model for COVID-19 patients using cough information is discussed by Pahar et al. (2022). The PCA and ML algorithms are used for object recognition. The Fast-RCNN was implemented by Ren et al. (2015) to make the process fast and to run the model. The small objects were recognized by Liu et al. (2016) with the contribution of YOLO algorithm. The SSD algorithm was implemented by Ganiger et al. (2017) to have a control on the crowd management system. Kahale (2020) described the measures in social distancing and provided the information which can reduce economically. The tracking of person whether an individual is affected by COVID-19 is determined by the model proposed by Georgievski (2020). The tracking and object detection algorithms are proposed by Georgievski (Harakannanavar et al. 2019a). The various deep learning models and their impacts were addressed in Wang et al. (Harakannanavar et al. 2018a) model. The model for detection of an individual is carried with video frames and Gaussian mixture for deep learning approach. The CNN approach in order to detect the human is carried out by Brunetti et al. (Harakannanavar et al. 2019b). The control of static approach was discussed by the Manfredi et al. (Harakannanavar et al. 2020a). Here, the text features are extracted and the classification was carried out using SVM. Huang et al. (Harakannanavar et al. 2018b) cameras were installed to monitor and find out the individual in the crowd management system. The significant information with respect to D detectors was described in the model suggested by Zou (Harakannanavar et al. 2018c). The social distance monitoring and controlling were made by using the networks suggested by Landing et al. (Harakannanavar et al. 2019c). The NN which uses synthetic data was suggested by Deng et al. (Harakannanavar et al.

2019d) to propose stereo images. The fused 3D point clouds are described in the suggested model by Zhang et al. (Harakannanavar et al. 2019e) to obtain the outdoor scenes.

2.1 Open Issues and Challenges (Wang et al. 2017)

- One of the primary drawbacks of the existing system is that it is not user-friendly because the data retrieval is very slow, and data is not maintained efficiently (Brunetti et al. 2018).
- The existing systems are not completely automated meaning the system needs to be monitored (Manfredi and Calderara 2014).

The systems raise a lot of privacy concerns that have been neglected until now. The system needs to be real-time. It fails to give high-end detail in surveillance that can be neglected through the naked eye (Huang et al. 2010).

2.2 Problem Definition

A surveillance system for monitoring the social distancing norms is needed to fight against COVID-19. To propose the proper solution for the above problem, a system that implements object detection on real-time video and monitors social distancing between human beings using a deep learning framework is required to develop. The model is explained in Sect. 3.

3 Proposed Methodology

The proposed system is validated using PyCharm (IDE) which integrates the laptop camera collecting real-time input. Initially, the video is captured frame-by-frame and resized to 300 $\times$ 300 pixels to meet the object detection model requirements. This is done with the help of the OpenCV module. SSD Mobile Net CAFFE model is a pre-trained deep learning model, used for object detection. Here, the resized frames are fed to the system. The model has been pre-trained for 20 labels such as bottle, bicycle, chair, monitor. For the model to detect specifically humans, conditional statements are used that filter only the human class. The class object humans are assigned a class id of 15. The humans are detected by drawing a bounding box around them. Then, calculate the mid-point by determining x and y coordinates of the bounding boxes. The mid-point helps us to find the distance between two bounding boxes or people. The model then enters a conditional loop, which determines whether the subjects (humans) are maintaining 2 m or not. If they are maintaining 2 m or more, the color

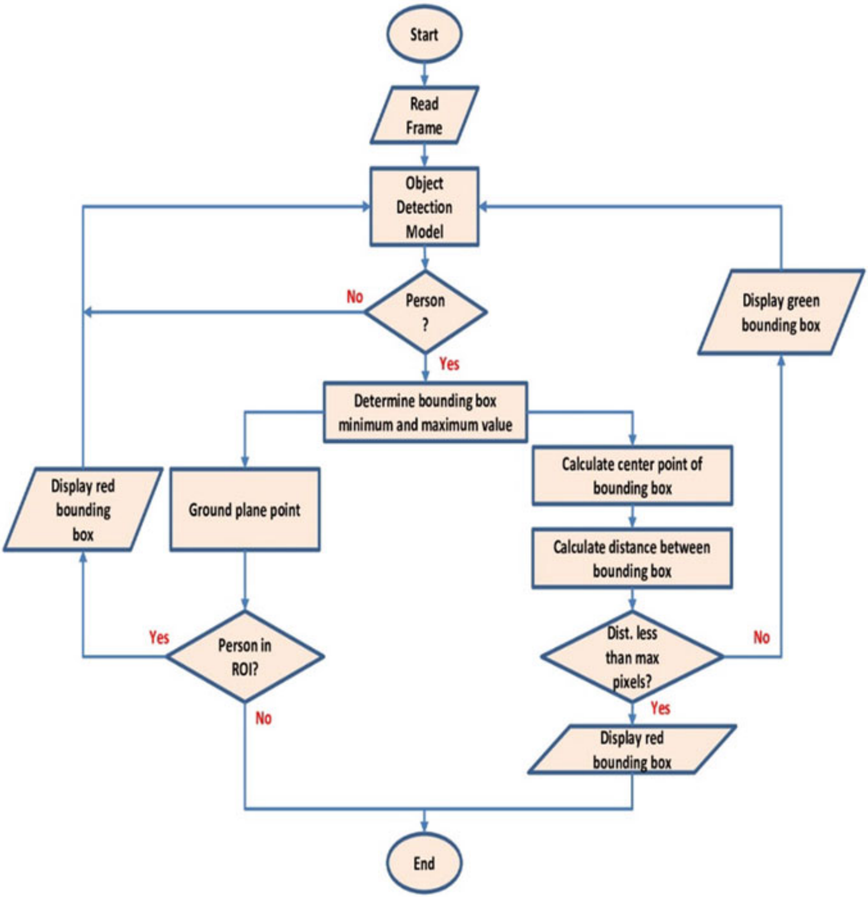

Fig. 1 Proposed model

of the bounding box appears green, and the non-alarming visual cue signifies safety and social distancing norms are met. If the subjects are maintaining less than 2 m, the color of the bounding box appears red, and the non-alarming visual cue indicates threat of exposure toward COVID-19 and is expected to maintain social distance. The proposed model is shown in Fig. 1.

3.1 Dataset Description

The proposed deep learning model has been pre-trained and tested on MS COCO dataset (Zou et al. 2019). Microsoft Common Objects in Context is widely known for its 328,000 organized and labeled images. It contains images of other things or stuff such as chair, grass, table, bicycle, airplane, street, background. More the number of images in the dataset, more accurately the deep learning model will learn to identify and distinguish between objects (human being). It is open source and suitable for implementation in machine learning projects (Landing 2020).

3.2 Finding the Mid-point of the Co-ordinates

Once the bounding box is created around a person in the frame, the system computes the four coordinates. Using the mid-point formula, then find the center point (Zhang and Vetter 2015). This center point will now act as a reference for other bounding boxes in the frame. The mid-point formula is $C\,(x, y) = (x_{\min} + x_{\max}/2, y_{\min} + y_{\max}/2)$.

3.3 Finding the Distance Between Two Bounding Boxes

$C_1(x_{\min}, y_{\min})$ and $C_2(x_{\max}, y_{\max})$ are the four coordinates that were determined in the previous process. The distance between each of the detected people in the frame is measured by evaluating the distance between their respective mid-points (Harakannanavar and Puranikmath 2017). The distance is calculated using Eq. 1.

$$d(C_1, C_2) = \sqrt{(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2}. \tag{1}$$

3.4 Modules

Modules can be defined as units of the project that work in tandem to ensure the smooth running of the system. The three key modules are as follows (Harakannanavar et al. 2019f):

- Input module.
- Human detection module.
- Recognition module.

3.5 Input Module

Here, two options are provided in view of input for the proposed system. Own laptop recorded with real time video. In case of option 2, using a pre-recorded video file for

Fig. 2 Input module example 1

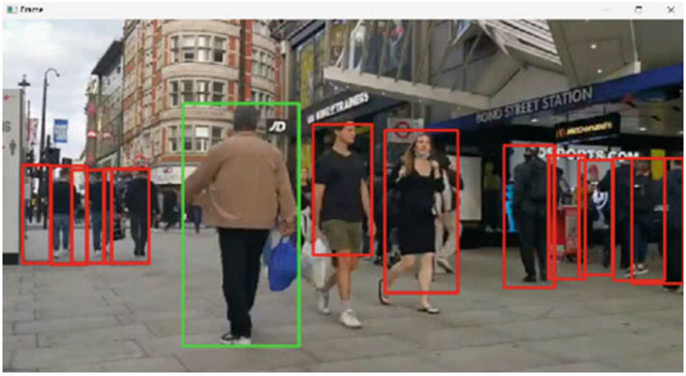

Fig. 3 Human detection module

which the file path will be given as input in the code and process further to produce the desired output. The input is collected and calibrated frame-by-frame using feature extraction. Now, import the OpenCV module from the Python library into virtual environment. It is responsible for image processing and development of real-time computer vision applications (Harakannanavar et al. 2019g). The input module and human detection module are shown in Figs. 2, 3, and 4, respectively.

3.6 Human Detection Module

The SSD Mobile Net V2 is an object detection module that was considered in the proposed model. It is pre-trained on the Common Objects in Context dataset which consists of 164K images (Harakannanavar et al. 2020b). Once the real-time video starts running, the model captures multiple frames successively and detects human

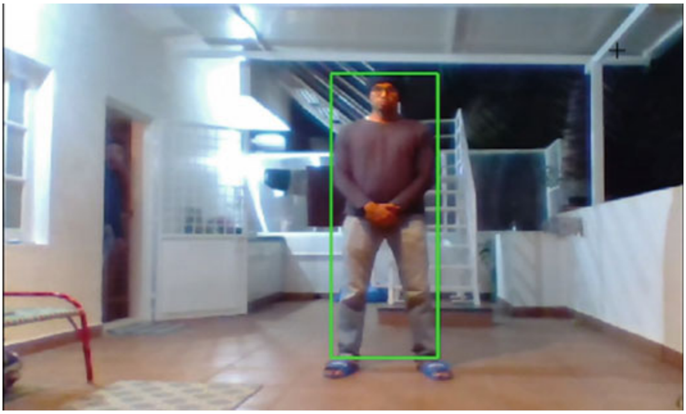

Fig. 4 Result Example 1

according to the assigned class object. Here, open-source software library is used to improve on the object class detection.

For training the deep learning module, the Common Objects in Context dataset also known as MS COCO (Harakannanavar et al. 2022a) is used for obtaining the results. Next, assign a class id to objects in the frame that categorize into humans. Although the model can identify a vast number of objects, that were focusing only identify the person in the frame with accuracy (Jayalaxmi et al. 2022). However, due to camera limitations, the accuracy reduces as the subjects move far from the camera (Kajabad 2019). The proposed model is tested for other object classes such as cat, dog, motorcycle, pots, TV, sofa, and monitor (Kreiss 2020).

4 Result Analysis

Following are the software requirements used PyCharm and Python. The hardware requirements are Laptop Processor (minimum) (Harakannanavar et al. 2022b): Pentium 4 (2.4 GHz), Hard Disk space (minimum): 250 GB, Random Access Memory (minimum): 1 GB, and camera and resolution: HD webcam (1280 × 720p) which are used to conduct the experiments. The video snippets below show the results of the human detection module. In this work, the video is recorded in an open space with dimensions 5.5 m depth × 3 m width at a fixed angle to replicate a surveillance system. Own laptop camera (Harakannanavar et al. 2022b) is used to capture the images for the implementation of the module, and the perspective view of the video frame is transformed to a bench view approximately 0.5 m above the ground. Figure 4 shows successful human detection along with a bounding box. The detection is a complete top-down detection, and since there is only a single person in the frame, the box will appear in green.

Next, another person in the frame to test the modules accuracy is considered. The model primarily focuses on measuring the distance between two or more people and emitting a non-alarming visual cue. Here, red indicates that the subjects in the frame are not maintaining social distancing norms. As mentioned, the permissible distance between two people is set to be 2 m, and the bounding boxes will appear green if the distance is greater than the threshold. In Fig. 5, the above-discussed framework is successfully employed. The neighboring bounding boxes appear in red color.

Next, the subjects in the frame maintain social distancing measures and the system being real-time tracks their movement and calculates the distance. The BB immediately changes to green as soon as the subjects in the frame move beyond threshold. A model is developed that would be applicable in social situations and tested the module for three people (Fig. 6).

Fig. 5 Result Example 2

Fig. 6 Result Example 3

In Fig. 7, third team member is positioned between the two. Since the three members are not following the protocol in respect to each other, all boxes over here appear red.

Next step is to make everyone in the frame to be positioned at 2 m distance from each other. The three bounding boxes appear green and is shown in Fig. 8.

Now, another permutation of positioning the three subjects in the frame to imitate a real-world scenario where there are constant movements is considered for the next event. The model works for all sorts of positioning. Figure 9 shows two members who are not following the social distancing measures in respect to a third person who is standing at the far end to the right. The visual cue alters according to the scenario indicating a real-time efficiency of the model.

Another variation in the positioning is carried out further and is shown in Fig. 10. It is observed that fourth member entered the frame who was maintaining distance in respect to all. The person on the right is following social distancing too.

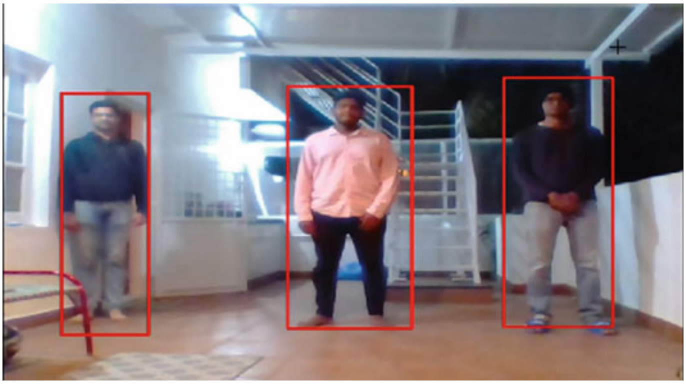

Fig. 7 Result Example 4

Fig. 8 Result Example 5

Fig. 9 Result Example 6

Fig. 10 Result Example 7

For verifying the efficiency of the code, rest of the two members were closely positioned behind. The model accurately calculates the depth and lateral distance between the all the members.

5 Conclusion and Future Scope

Social detection and distancing monitoring using a DL model are implemented by taking the frames of an input video or from a real-time video recording. The DL model is deployed to detect humans and create a bounding box around them. The spatial coordinates and the mid-point of the bounding box are calculated. The proposed model using deep learning is tested on a video portraying pedestrian on a street to identify, and the model can be used as a solid foundation for the development of projects in the field of study such as, study of behavioral patterns of people, predicting crowd or traffic movement. The model proved with an accuracy of 98.6%

on real-time dataset. In future, the model can be modified and customized for use in other environments such as shopping malls, cinema theaters, offices, restaurants, and schools.

References

Ahmed I, Ahmad M, Rodrigues JJ, Jeon G, Din S (2021) A deep learning-based social distance monitoring framework for COVID 19. Int J Sustain Cities Soc 65:1–12

Brunetti A, Buongiorno D, Trotta GF, Bevilacqua V (2018) Neuro computing computer vision and deep learning techniques for pedestrian detection and tracking: a survey. J Neurocomputing 300:17–33

Ganiger D, Patil KA, Patil P, Anandhalli M (2017) Automatic control of power supply in classroom using communication technology. In: IEEE ınternaional conderence of applications of electronics and communication technologies, pp 230–234

Georgievski (2020) Tracking COVID-19. Biomedical Eng Community 1–4

Gopal B, Ganesan A (2022) Real time deep learning framework to monitor social distancing using improved single shot detector based on overhead position. Earth Science Inform 585–602

Harakannanavar SS, Jayalaxmi H, Asha CN, Prashanth K, Hudedavar P (2022a) Biometric trait: offline signature ıdentification and verification based on multimodal fusion techniques. J Posit Sch Psychol 6(4):2180–2191

Harakannanavar SS, Ramachandra AC, Pramodhini R, Surekha M, Puranikmath VI, Prashanth CR (2022b) Performance evaluation of feature level fusion for multimodal biometric systems. Math Stat Eng Appl 71(4):2775–2792

Harakannanavar SS, Asharani TS (2018) Design of an efficient algorithm for Iris recognition. In: IEEE international conference on inventive communication and computational technologies, pp 876–880. https://doi.org/10.1109/ICICCT.2018.8473179

Harakannanavar SS, Puranikmath VI (2017) Comparative survey of ıris recognition. In: IEEE ınternational conference on electrical, electronics, communication, computer, and optimization techniques, pp 280–283

Harakannanavar SS, Prashanth CR, Raja KB, Chetan M (2018a) Face recognition based on the fusion of bit-plane and binary ımage compression techniques. In: IEEE ınternational conference on recent trends in electronics, ınformation and communication technology, pp 1889–1994

Harakannanavar SS, Vidyashree, Prashanth CR, Raja KB (2018b) Performance evaluation of face recognition based on multiple feature descriptors using Euclidean distance classifier. Int J Adv Netw Appl 10(3):3864–3879

Harakannanavar SS, Prashanth CR, Raja KB (2018c) Performance evaluation of face recognition based on the fusion of bit-plane and binary ımage compression techniques using euclidean distance classifier. Int J Intell Eng Syst 11(6):52–64

Harakannanavar SS, Prashanth CR, Raja KB, Patil S (2019a) Face recognition based on SWT, DCT and LTP. In: International conference on ıntegrated ıntelligent computing, communication and security, pp 565–573

Harakannanavar SS, Prashanth CR, Raja KB (2019b) MSB based ıris recognition using multiple feature descriptors. In: International conference on ınternational conference on ıntelligent computing, ınformation and control systems, pp 615–623

Harakannanavar SS, Vidyashree, Puranikmath VI, Prashanth CR, Raja KB (2019c) Comprehensive study of biometric authentication systems, challenges and future trends. Int J Adv Netw Appl 10(4):3958–3968

Harakannanavar SS, Vidyashree, Puranikmath VI, Prashanth CR, Raja KB (2019d) An extensive study of ıssues, challenges and achievements in ıris recognition. Asian J Electr Sci 8(1):25–35

Harakannanavar SS, Vidyashree, Puranikmath VI, Prashanth CR, Raja KB (2019e) Technical challenges, performance metrics and advancements in face recognition system. Int J Comput Sci Technol 7(3):836–847

Harakannanavar SS, Prashanth CR, Raja KB (2019f) Performance analysis of MSB based ıris recognition using hybrid features extraction technique. Int J Recent Technol Eng 8(6):1–10

Harakannanavar SS, Prashanth CR, Raja KB (2019g) Performance analysis of ıris recognition using multi-stage wavelet transform decomposition and bicubic ınterpolation technique. Int J Recent Technol Eng 8(6):230–239

Harakannanavar SS, Prashanth CR, Raja KB (2020a) Iris recognition using bicubic ınterpolation and multi level DWT decomposition. In: International conference on computational vision and bio ınspired computing, pp 1146–1153

Harakannanavar SS, Rudagi J, Puranikmath V (2020b) Development of algorithm for offline signature verification using fusion extraction techniques. Int J Seybold Rep 15(9):1421–1431

Huang P, Hilton A, Starck J (2010) Shape similarity for 3d video sequences of people. Int J Comput Vision 89(2):362–381

Jayalaxmi H, Anitha TG, Harakannanavar SS, Prashanthi HJ (2022) Robust ıris recognition algorithm using EMD and support vector machine. J Posit Sch Psychol 6(4):4279–4288

Kahale (2020) On the economic impact of social distancing measures. SSRN Electron J 89–98

Kajabad I (2019) People detection and finding attractive areas by the use of movement detection analysis and deep learning approach. Procedia Comput Sci 156:327–337

Kreiss A (2020) Joint human pose estimation and stereo 3D localization. In: IEEE ınternational conference on robotics and automation, pp 2324–2330

Landing (2020) A.I. Landing AI Creates an AI Tool to Help Customers Monitor Social Distancing in the Workplace. IEEE Magazines

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) SSD: single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37

Manfredi V, Calderara C (2014) Detection of static groups and crowds gathered in open spaces by texture classification. Pattern Recogn Lett 44:39–48

Mangshor NN, Rajudin MN, Aminuddin R (2021) Framework for JARAK: a monitoring and notifier system for social distancing. In: IEEE student conference on research and development, pp 422–426

Pahar M, Niesler T (2022) Machine learning based COVID-19 detection from smartphone recordings: cough, breath and speech. arXiv e-prints, arXiv-2104

Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detectionwith region proposal networks. arXiv preprint arXiv:1506.01497

Wang X, Ng W, Liang (2017) Lapped convolutional neural nettworks for embedded systems. In: IEEE global conference on signal and ınformation processing, pp 1135–1139

Yadav S, Gulia P, Gill NS, Chatterjee JM (2022) A real-time crowd monitoring and management system for social distance classification and healthcare using deep learning. Hindawi J Healthc Eng 1–11

Zhang C, Vetter Z (2015) Sensor fusion for semantic segmentation of urban scenes. In: IEEE ınternational conference on robotics and automation, pp 1850–1857

Zou S, Guo Y, Ye J (2019) Object detection in 20 years: a survey. arXiv