ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse1211022021.pdf
https://doi.org/10.30534/ijatcse/2021/1211022021

Implementing a Real Time Human Detection and Monitoring Social Distancing for Covid-19 Using V-J algorithm and OpenCV

Raja Rajeshwari Jadhav[1], Dr. S. L. Lahudkar[2]
[1] JSPM's ICOER, Wagholi, Savitribai Phule Pune University, Pune, rajrajeshwari.m@gmail.com
[2] JSPM's ICOER, Wagholi, Savitribai Phule Pune University, Pune, sllahudkar_entc@jspmicoer.edu.in

ABSTRACT

Recently, the outbreak of Coronavirus Disease (COVID-19) has spread rapidly across the world and thus social distancing has become one of mandatory preventive measures to avoid physical contact. COVID-19 is a disease caused by a severe respiratory syndrome coronavirus. The continuous development of technology of IT enabled computers to see and learn. There are many viable applications for computer learning and vision to solve new tasks. In this paper, we propose a framework, able of automatically detect no. of human bodies present in a single image, acquired by a traditional low-cost camera. In this paper Viola jones algorithm is used to detect human monitoring social distancing norm. System is divided into two parts, the first part is about person detection whereas second part is about monitoring whether people are following social distancing or not, it is applicable if image contains more than one human. This paper is going to study and understand the Viola-Jones algorithm by implementing the whole detection framework and based on the implementation, conduct experiment to hopefully further improve the performance.

Key words : Adaboost, Cascade classifier, Haar-like features, Social Distancing norm, Viola jones.

1. INTRODUCTION

COVID-19 has spread rapidly worldwide since it began, greatly affecting peoples' lives, social economies and medical systems. At present, little is known about the disease, and vaccines are still under development. Therefore, in the face of severe outbreaks, previous effective experience can help people better protect themselves and their families. The aim of this article is to discuss the social distancing measures for COVID-19[1].

Social distancing involved keeping a distance of 1.5 m between people, which can prevent the spread of most respiratory infectious diseases. Social distancing is one of the most effective measures to reduce the spread of the virus, which is transmitted by air droplets[2]. The droplets produced by coughing, sneezing or forced speaking have a certain transmission distance. By keeping this distance, we can reduce the spread of the virus. Wearing masks, washing hands frequently and disinfecting with alcohol also help to prevent the virus from spreading from one person to another. To control the disease, the World Health Organization recommended that countries should strengthen case detection, track and monitor contacts, practice isolation from close contacts and isolate cases as well as implement traffic control and suspend large gatherings.

2. LITERATURE SURVEY

There are many security techniques are existed for the human detection and social distancing. For the implementation of work some papers are studied which are listed below.

The survey paper, "Monitoring Social Distancing for Covid-19 Using OpenCV and Deep Learning", emphasizes on a surveillance method which uses Open-CV, Computer vision and deep learning to keep a track on the pedestrians and avoid overcrowding. The implementation can be done using closed circuit television (CCTV) and Drones where the camera will detect the crowd with the help of object detection and compute the distance between them. The Euclidean distance between two people will be calculated in pixels and is compared with given standard distance and if it is observed to be less than the standard distances the local authorities or local police authorities will be notified [2].

In [3], Bharathi B and Bhuvana J have proposed an automatic human activity recognition system that independently recognizes the actions of the humans. The results obtained show that, the Multilayer Perceptron has obtained 98.46% of overall accuracy in detecting the activities. The second-best performance was observed when the classifiers are combined together.

In [4], Ahmad al-Qerem and Arwa Alahmad proposed a framework, able of automatically perceiving the human body poses from a single image, acquired by a traditional low-cost

camera. The processing start with detection human in image and then extracting the silhouette from an image then using a neural network to recognize body poses based on silhouettes that extracted.

It is widely recognized that video processing and object detection are computing intensive and too expensive to be handled by resourcelimited edge devices. Inspired by the depthwise separable convolution and Single Shot Multi-Box Detector (SSD), a lightweight Convolutional Neural Network (L-CNN) is introduced in [5]

In [6], Jong Hyun Kim et al studied a CNN-based method for human detection in a variety of environments using a single image captured from a visible light camera at night. When we compared the detection accuracies of a CNN using pre-processing by HE and a CNN that did not use this pre-processing, we found that the accuracy was higher when using a combined database with HE pre-processing.

M. Kumar and P. Torr, [7]. In Exemplar based method, an exemplar pool should be constructed first, and then, test images are matched with the exemplars or models. They develop an efficient method, OBJCUT, to obtain segmentations using our probabilistic framework. Drawback of this method is that, it cannot always accurately segment body, because exemplars cannot cover all situations of poses and appearance variation.

Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, [8]. They proposed a method that incorporates local MRF and global shape priors to iteratively estimate segmentations and pose simultaneously.

P. Kohli, J. Rihan, M. Bray, and P. Torr, [9]. They use pose specific CRF for segmentation and pose estimation which has been successfully used for 3-D Human pose tracking. The experimental results show that this model suffices to ensure human-like segmentations.

X. Ren, A. C. Berg, and J. Malik, [10]. They developed a framework that uses arbitrary pairwise constraints between body parts. Probability of boundary (Pb)can be used as part detectors and then by assembling the candidates with integer quadratic programming human body configurations are recovered. It is very hard to design a robust part detector and its performance is also limited.

V. Ferrari, M. Marín-Jiménez, and A. Zisserman, [11]. Ferrari et al. use GrabCut algorithm to segment human based on head detector. In which first two stages i.e. Human detection and tracking and Foreground highlighting use a weak model of a person obtained through an upper-body detector generic over pose and appearance. This weak model only determines the approximate location and scale of the person, and roughly where the torso and head should lie. However, it knows nothing about the arms, and therefore very little about pose.

L. Bourdev, S. Maji, T. Brox, and J. Malik, [12]. Bourdev et al. utilized a poselet detector to detect keypoints of human. The trained poselet masks are combined with boundary cues to detect specific poselet and finally the human is segmented.

3. PROPOSED SYSTEM

A recent study indicates that social distancing is an important containment measure and essential to prevent SARSCoV-2, because people with mild or no symptoms may fortuitously carry corona infection and can infect others[1]. This study aims to support the reduction of the corona virus spread and its economic costs by providing an AI-based solution to automatically monitor and detect violations of social distancing among individuals. Block diagram of proposed system is shown in figure 1.

Image from surveillance camera or from database is taken as input. Pre-processing eliminates background; de-noise the image and cuts out unimportant parts from image. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions (i.e. noise removing) or enhances some image features important for further processing.

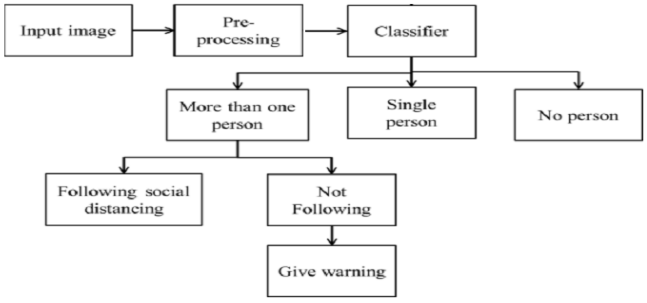

Figure 1: Architecture of Proposed System

In figure 1, the main goal is to identify which class/category the new data will fall into. Classification is performed using Viola Jones algorithm (V-J). Input to classifier is real time image from a camera. Classifier first detects the face of a human from input images and thus detects human in image. It then classifies image into following category

- No person detected: social distancing is invalid
- Single person is detected: no need of social distancing
- Multiple people detected: social distancing is necessary. System checks whether every person in frame is obeying social distancing norm or distance between them is 2 meter or not, if not them system gives warning.

4. VIOLA JONES ALGORITHM

One important Computer Vision application is the ability to have a computer detects objects in images. Among those objects, the human face receives the most attention since it has many useful applications in security and entertainments. Developed in 2001 by Paul Viola and Michael Jones, the Viola-Jones algorithm is an object-recognition framework that allows the detection of image features in real-time.Viola-Jones is quite powerful and its application has proven to be exceptionally notable in real-time face detection[13].

The Viola-Jones algorithm consists of 4 main steps which are mainly-

1. Selecting Haar-like features
2. Creating an Integral image
3. Running AdaBoost training
4. Creating classifier cascades

1.1 Selecting Haar-like features

Haar-like features are named after Alfred Haar, a Hungarian mathematician in the 19th century who developed the concept of Haar wavelets (kind of like the ancestor of haar-like features). The features below show a box with a light side and a dark side, which is how the machine determines what the feature is. Sometimes one side will be lighter than the other, as in an edge of an eyebrow. Sometimes the middle portion may be shinier than the surrounding boxes, which can be interpreted as a nose[15].

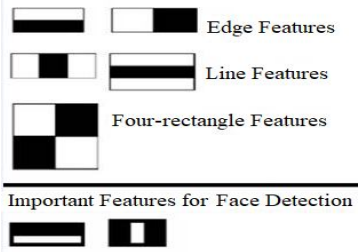

Figure 2 : Haar-like features

There are 3 types of Haar-like features that Viola and Jones identified in their research:

- Edge features
- Line-features
- Four-sided features

A Haar-like feature consists of dark regions and light regions as shown in figure 2. It produces a single value by taking the sum of the intensities of the light regions and subtract that by the sum of the intensities of dark regions.Now, when feature window moves over the eyes, it will compute a single value which will be compared to some threshold and if it crosses the threshold it will conclude the existence of an edge or a positive feature. Using a variety of different combinations of these features we can determine whether an image contains a human face or not. As mentioned, the Viola-Jones algorithm calculates a lot of these features in many subregions of an image which becomes computationally expensive. To tackle this problem, Viola and Jones used concept of Integral Images[17].

1.2 Creating an integral image

The framework devised by Viola-Jones uses a base window of size 24X24, which leads to over more than 180,000 features being calculated in this window. It becomes very computationally expensive to calculate the pixel difference for all the features. To avoid such costly computations the concept of an integral image (also known as a summed-area table) was introduced. It uses a quick and efficient way to calculate the sum of pixel values in an image or rectangular part of an image. In the integral image, value at each point is determined by summing all the pixels which are above and the pixels which are to the left and also the target pixel itself as shown in figure 3. The integral image can be calculated in a single pass over the original image.

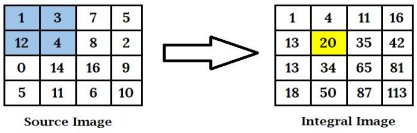

Figure 3 : source image and integral image

To determine the sum of pixels under any given rectangle, we use the integral image and require only the 4 corner values instead of summing all underlying pixels individually. So, there is a significant reduction in computation as now we only need 3 operations and 4 corner values, regardless of rectangle size[16].

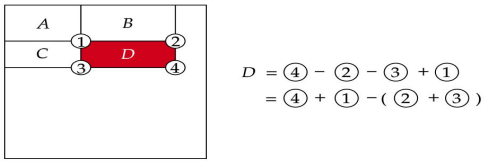

Figure 4 : calculating a rectangle region using integral image

Intermediate representation is essential because it allows for fast calculation of rectangular region. To illustrate, Figure 4 shows that the sum of the red region D can be calculated in constant time instead of having to loop through all the pixels in that region. Since the process of extracting Haar-like features involves calculating the sum of dark/light rectangular regions, the introduction of Integral Images greatly cuts down the time needed to complete this task.

1.3 Adaboost

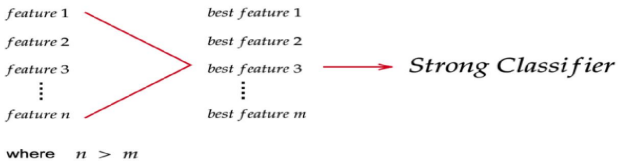

Figure 5 : The goal of using the AdaBoost algorithm is to extract the best features from n features. Note: best features are also known as weak classifiers.

The AdaBoost (Adaptive Boosting) Algorithm is a machine learning algorithm for selecting the best subset of features among all available features. From figure 5, the output of the

algorithm is a classifier (a.k.a Prediction Function, Hypothesis Function) called a "Strong Classifier". A Strong Classifier is made up of a linear combinations of "Weak Classifiers" (best features). From a high level, in order to find these weak classifiers the algorithm runs for T iterations where T is the number of weak classifiers to find and it is set by you. In each iteration, the algorithm finds the error rate for all features and then choose the feature with the lowest error rate for that iteration[14].

1.4 Cascading Classifiers

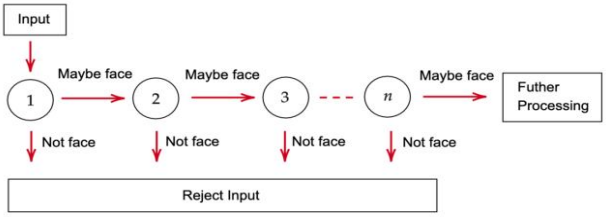

Figure 6 : Cascade Classifier

A Cascade Classifier is a multi-stage classifier that can perform detection quickly and accurately. In figure 6, each stage consists of a strong classifier produced by the AdaBoost Algorithm. From one stage to another, the number of weak classifiers in a strong classifier increases. An input is evaluated on a sequential (stage by stage) basis. If a classifier for a specific stage outputs a negative result, the input is discarded immediately. In case the output is positive, the input is forwarded onto the next stage. According to Viola & Jones (2001), this multi-stage approach allows for the construction of simpler classifiers which can then be used to reject most negative (non face) input quickly while spending more time on positive (face) input[13].

5. RESULT

While running the project, the camera is started and it takes the input image from the camera. Social distance monitoring is not valid for No person or 1 person detected. If the input image has no person, then the output will display "person not detected". If the input image has only one human, then the ouput will be "1 person detected".

Figure 7 : multiple people detected and social distancing maintained.

In figure 7,more than one person is detected and our proposed system checks whether people are following social distancing or not, it is applicable if image contains more than one human. Here people are following social distancing and message will be display "all okay" and "more than 1 person detected".

Figure 8 : social distancing norm is not followed

In figure 8, multiple person is detected but social distancing norm is not followed. So voice message gets displayed as "Please maintain social distancing" and "more than 1 person detected".

Table 1. Testing Results Face Counts from Images

No	Image	Count face an image		Accuracy	Processing Time(second)
		Success	Failure		
1	Image 1	8	-	100%	10.2
2	Image 2	6	6	100%	2.7
3	Image 3	4	-	100%	4.2
4	Image 4	5	-	100%	3.1
5	Image 5	11	-	100%	4.2
6	Image 6	24	3	88.88%	12.2
7	Image 7	30	1	93.70%	13.6
8	Image 8	3	-	100%	3.6
9	Image 9	2	-	100%	2.4
10	Image 10	8	2	80%	14,5
11	Image 11	12	-	100%	3.7
12	Image 12	36	2	94%	5.9
13	Image 13	5	-	100%	3.3
14	Image 14	4	-	100%	2.4
15	Image 15	12	-	100%	3.7
16	Image 16	8	-	100%	2.9

No	Image	Count face an image		Accuracy	Processing Time(second)
		Success	Failure		
17	Image 17	5	5	100%	2.3
18	Image 18	24	3	88.88%	4.5
19	Image 19	3	-	100%	2.3
20	Image 20	24	1	96%	4.9
21	Image 21	4	-	100%	1.4
22	Image 22	-	-	0%	1.2
23	Image 23	5	-	100%	1.5
24	Image 24	1	-	100%	2.4
25	Image 25	3	-	100%	7
26	Image 26	1	-	100%	1.3
27	Image 27	-	-	0%	1.2
28	Image 28	5	-	100%	1.4
29	Image 29	1	-	100%	22.2
30	Image 30	3	-	100%	3.3
31	Image 31	1	-	100%	1.8
32	Image 32	1	-	100%	1.4
33	Image 33	5	-	100%	2.2
34	Image 34	3		100%	2
35	Image 35	10	-	100%	11.5
36	Image 36	50	1	98%	31.5
37	Image 37	2	-	100%	2.1
38	Image 38	31	1	96.80%	13.3
39	Image 39	1	-	100%	1.7
40	Image 40	28	2	93.30%	4.9
Accuracy = 93.24%					

From table 1 testing the face count using the viola jones method that has been done successfully detected the number of faces in the image as much as 93.24%. One of the causes of failure to detect faces is that the light in the image is too dark so the number of faces displayed is incorrect.

6. CONCLUSION

The proposed system can detect humans automatically and does not need a manual detection threshold to select one that has the highest true positive rate. Accuracy of proposed system is 93.24%. The proposed system achieves real-time performance for testing live-captured videos because the optical flow model only utilizes two successive frames to find motion. Moreover, deep models only require a single frame to classify the optical flow patches as human or nonhuman (i.e., human detection). The drawback of the proposed system is that it is highly dependent on the quality of the optical flow processing stage. Adding tracking to the whole pipeline for human detection can reduce this dependency and improve the overall accuracy. For a future work, we will integrate tracking that makes use of initially extracted training regions around humans as positive samples and other regions as negative samples. As the result is highly accurate and efficient, we will utilize the results of human detection demonstrated in this paper for human action recognition to map each activity with a specific action class.

REFERENCES

1. Nouar AlDahoul , Aznul Qalid Md Sabri , and Ali Mohammed Mansoor **Real-Time Human Detection for Aerial Captured Video Sequences via Deep Models** Vol. 2018, Article ID 1639561, pp. 1-14, Feb. 2018
2. Rucha Visal, Atharva Theurkar, Bhairavi Shukla, **Monitoring Social Distancing for Covid-19 Using OpenCV and Deep Learning**, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 | June 2020 Page 2258
3. B, Bhuvana J **Human Activity Recognition using Deep and Machine Learning Algorithms** International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-4, February 2020
4. Ahmad al-Qerem1 Arwa Alahmad **Human Body Poses Recognition Using Neural Networks with Data Augmentation** International Journal of Advanced Trends in Computer Science and Engineering, 8(5),September - October 2019, 2117 – 2120
5. Seyed Yahya Nikouei, Yu Chen, Sejun Song, Ronghua Xu, Baek-Young Choi, Timothy R. Faughnan, **Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN**, arXiv:1805.00330v1 [cs.CV] 24 Apr 2018
6. Jong Hyun Kim, Hyung Gil Hong and Kang Ryoung Park, **Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light**

Camera Sensors, Sensors 2017, 17, 1065; doi:10.3390/s17051065

7. M. Kumar and P. Torr, **OBJCUT: Efficient segmentation using top-down and bottom-up cues**, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 3, pp. 530–545, Mar. 2010.
8. Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, **An interactive approach to pose-assisted and appearance-based segmentation of human** ,in Proc. ICCV, 2007, pp. 1–8.
9. P. Kohli, J. Rihan, M. Bray, and P. Torr, **Simultaneous segmentation and pose estimation of humans using dynamic graph cut**, Int. J. Comput. Vis., vol. 79, no. 3, pp. 285–298, Sep. 2008.
10. X. Ren, A. C. Berg, and J. Malik, **Recovering human body configurations using pairwise constraints between parts**, in Proc. ICCV, 2005,pp. 824–831.
11. V. Ferrari, M. Marín-Jiménez, and A. Zisserman, **2D human pose estimation in TV shows, Statistical and Geometrical Approaches to Visual Motion Analysis**, pp. 128–147, 2009.
12. L. Bourdev, S. Maji, T. Brox, and J. Malik, "**Detecting people using mutually consistent poselet activations**," in Proc. ECCV, 2010,pp. 168–181.
13.] P. Viola and M. Jones, **Rapid object detection using a boosted cascade of simple features**, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, vol. 1, pp. I-511-I–518.
14. S. pandey, **Review and Comparison of Face Detection Algorithms**, Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 4111–4117, 2014.
15. Xiaowei Zhao, XiujuanChai, **Context Constrained FacialLandmark Localization Based on Discontinuous Haar-like Feature** International Conference on Computer Vision (ICCV2013),2013.
16. Cordiner, A.; Ogunbona, P.; Wanqing Li, **Face detection using generalised integral image features**, Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.1229,1232, 7- 10 Nov. 2009
17. Lienhart and J. Maydt. **An Extended Set of Haar-like Features for Rapid Object Detection**, IEEE ICIP 2002