An Approach for Person Detection along with Object Using Machine Learning

Srikanth Bethu [1,*], M. Neelakantappa [2], A. Swami Goud [3], B. Hari Krishna [4], and P. N. V. Syamala Rao M [5]

[1] Department of CSE, CVR College of Engineering, Hyderabad, Telangana, India
[2] Department of IT, Vasavi College of Engineering, Hyderabad, Telangana, India; Email: m.neelakanta@gmail.com (M.N.)
[3] Department of IT, B V Raju Institute of Technology, Narsapur, Telangana, India; Email: swamygoud.a@bvrit.ac.in (A.S.G.)
[4] Department of CSE, Mallareddy Engineering College, Hyderabad, Telangana, India; Email: haribоorgadda@gmail.com (B.H.K.)
[5] Department of CSE-AIML, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telangana, India; Email: syamalarao_p@vnrvjiet.in (P.N.V.S.R.M.)
*Correspondence: srikanthbethu@gmail.com (S.B.)

***Abstract*—The best biometric information processes is a face recognition device, its applicability is simpler and its working range is broader than other methods like fingerprinting, iris scanning and signature. Face Detection is one of the kinds of bio-metric strategies that immediately apply to facial recognition by computerized devices through staring at the facial. It is a common feature used in bio analytics, digital cameras, and social labeling. Main applications of facial recognition algorithms that concentrate on recognition of face include environments, artifacts, and other parts of humans. Face-detection systems uses learning algorithms which are part of machine learning that can be used to identify subject faces inside big size pictures in order to function.**

***Keywords*—face detection, human detection, machine learning, object detection**

I. Introduction

Face detection is one of the emerging technologies that can solve various social issues and can be applied in many fields such as medical, industrial, manufacturing, surveillance etc., to make work easier and more comfortable. Our human faces have lot of distinguishable features, we have lots of peaks and valleys in faces that help us to identify uniquely. Visonic says that these things are landmarks called the nodal points. Our human face has a maximum of 80 nodal points. Facial regions such as eyebrows, nostrils, eyes, mouth etc. are detected to identify a person. Locating a human face and comparing with the images in the database to identify the individual. This is done by using open Computer Vision (CV) where such ideas are made into reality. Our objective is to develop, a facial recognition system that has become vital in large cities. Many scientists and researchers are focusing on establishing robust and efficient algorithms for the detection of littering. The face of the person is detected and stored. The face of an individual is composed of different structures and Face recognition system is one of the efficient techniques and the system does not need any intervention of people to operate, that makes it possible to detect people from images obtained from the camera. There are many computer vision techniques proposed to deal with face detection or tasks with high robustness as well as discrimination. However, several issues still need to be addressed owing to various challenges, like head orientation, movements, littering while travelling as well as facial expression. The motivating techniques are developed to face all challenges, and thus develop trustworthy face recognition systems.

Detection of face can be considered as a special object-class detection event. When detecting object-type, the job is to figure out the positions and measurements of all objects in a picture belonging to a class which is given. Examples cover upper torsos, pedestrians, and motor vehicles. Detection of face algorithms concentrate on identifying human front side of a face. It is same as to that of image recognition, where a person's image is similar bit by bit. Picture suits the data base stores. Changes to the facial attribute in the database would invalidate the corresponding procedure. Here we are using Eigen face technique [1], we are first taking the data set and then we are training the dataset using some algorithms like Local Binary Pattern Histogram (LBPH) [2] and then recognizing the face. First, the potential human eye regions are identified in the gray-level picture by checking all the valley regions. The genetic algorithm is used to determine all possible areas that conflict with eyebrows, iris, and nostril and mouth corners.

Manuscript received September 13, 2022; revised November 11, 2022; accepted December 1, 2022; published May 5, 2023.

II. Related Work

A. Face Detection and Recognition Using OpenCV

The principal component analysis method to reduce the amount of storage which thereby gives rise to the projection of self-space. With open CV, face recognition can be done quickly. In this paper various types of face algorithms are studied and explained like Haar cascade, Eigen face, fisher faces and local binary patterns histograms. Here the detection algorithm Principal Component Analysis (PCA) [3] helps us to reduce large dimension into a smaller intrinsic dimension of the characteristic space. It is a linear domain technique used for linear models. The PCA depicts the large 1-D vector of pixels formed out of 2-D face in the main component of feature space. This is what we call the self-space projection. The process is like the noisy image undergoes pre-filtering, Wavelet Based Denoising-Principle Component Analysis (WBD-PCA) [4] is used as post-filtering, and finally we will get a reconstructed image.

B. Pedestrain Detection System Using Raspberry PI3

In the use of object detection to identify humans in the images is taken. Histogram of Gradients (HOG) and machine learning algorithms are used to identify the individual objects in images. They use open CV and MATLAB in raspberry pi3 board to implement it in real-time. The HOG extracts the features from the images and gives it for process to train Support Vector machine (SVM) with those pictures. HOG [5] will calculate the gradients of the image of size (x, y), the Red, Green and Blue (RGB) image is converted to gray scale by reducing the luminosity. It is then passed through the filters say F and F'. The support vector machine, SVM is a classifier which classifies the input image with different labels during the training stage. After classification, scaling of the image is done which is nothing but the image pyramid. Then sliding window technique is used to analyze the entire image with a predetermined window size. Non-maximum sup-pression algorithm is used to remove non-local maximum pixels around the selected pixel. NMS picks one image from the multiple images of one object. In raspberry pi3 an OS is installed followed by an open CV python setup with all packages. In training stage, we train it with two sets of images positive and negative set of images with and without pedestrians. Image set (positive and negative) → HOG → SVM (train → SVM (model) →SVM (predict) → NMS → Image detected. In detection stage it in volves the following process, Video → Frame conversion → image pyramid and slid- ing window → 64×128 image window → HOG.

C. Faster RCNN Detection Based on the OpenCV, CSRT Tracker Using Drone Data

Object tracking has various drawbacks like overlapping, camera motion blur, changing object appearance, environmental variation etc. Object classification is done from the Image frame followed by object detection, CSRT tracking and finally the output image we require. To overcome the drawbacks, they use CSRT tracker-based object detection. Object detection is done by faster Region-based Convolutional Neural Networks (R-CNN) [6] in live stream which is very much demanding for real life problems. Object tracking contains two processes tracking object classifier and generating object detection, b) implementation of tracking algorithm. Images are taken from drones, the images are placed in the convolution layer, then a feature map through a region proposal network with lots of proposals involves ROI pooling, classified and finally an output image is obtained. This is the faster R-CNN method. The output image is then sent for trained object detection model. The faster R-CNN is built with some help of Deep Neural Networks (DNN) layers where open CV DNN module can support easily. The tracking method involves frames converted through cv2.dnn.blob. From images to blob, then undergoes faster R-CNN object detection, output predictions of those detections and then Channel and Spatial Reliability Tracking (CSRT) tracker function CV Tracker CSRT where if it satisfies then tracking takes place else repeats from the faster RCNN object detection. Frame algorithm takes blobs and gives it for object detection to detect the location of object and classify object class. The output of the detection is sent to tracking algorithm to track the object specified, if any errors during the process the tracker will invoke the object and restart the process from first. The main goal of the paper is to design a simple object tracker that is reliable to the images captured by drones.

D. Multi-face Challenging Dataset for Robust Face Recognition

Face recognition of images in an active area. This is especially a challenging task. Due to variation of viewpoint, scale, pose, illumination, and expression, it becomes a challenge to the person to design the application. Dataset with pose variation, occlusion mask, spectacle etc. must be made to identify the person correctly. Face descriptors like Visual Geometry Group (VGG) Face (refers to the face recognition model made by the members of Visual Geometry group at the University of Oxford) [7] are used in this application.

III. Methodology

Data science is an inter-disciplinary field that employments logical strategies, forms, calculations and frameworks to extricate information and experiences from numerous structural and unstructured data. Data science is related to data mining and enormous information. Data science could be a "concept to bind together insights, data analysis, machine learning and their related strategies" in arrange to "get it and analyze real wonders" with data. It utilizes procedures and hypotheses drawn from numerous areas inside the setting of science, measurements, computer science, and data science.

Our objective is to create a framework that will utilize computer vision methods to consequently distinguish and recognize faces from the computerized pictures which are extricated from the input video. The recognizable proof

and acknowledgment is based on unmistakable facial highlights such as locale of the eyes, face shape etc. We are attempting to construct a quick and proficient confront acknowledgment system that recognizes faces exceptionally rapidly in cluttered foundations. Employing a learning-based approach, namely haar-cascade classifier, we need to play down the impacts of undesirable objects within the genuine time environment.

Using computer vision to detect someone missing the trash is an application but the thing is that the application does not identify the person nor label them or warn them so that such mistake can be avoided in the future. Moreover, if the application comes into play and will not allow litterer to do such thing as they will know that someone is watching them and if they do that mistake, they will be warned for that. This kind of catching the litterer during the incident and making them to understand far better than just identifying the trash been thrown. MIL trackers do a poor job of reporting failure whereas KCF does not handle full occlusion well and so we go for a better algorithm to use in the project called the CSRT algorithm (Discriminative Correlation filter with channel and spatial reliability) for a better tracking of the object.

In this paper we have a done face recognition system to identify people who litter in public. A face recognition system with a pre-determined dataset of a particular area of people is kept and if one among those people put their garbage in path rather than on dustbins, then those persons and identified and details are gathered. To identify the garbage been thrown we use HOG detector, which is usually used to determine pedestrians. Once it is detected a bounding box is created. The pixels within the bounding image were monitored, if the set of pixels crosses the boundary of the bounding box, then it was marked. The image is then stored, now a tracking algorithm is used to monitor the trash. CSRT [8] (Discriminative Correlation Filter with channel and spatial reliability) tracker is used. If the CSRT tracked the image move beyond the litterer, then the object will be considered as trash. This is how the litterer is identified and further procedures can be done.

Hog stands for Histogram of oriented gradients. It is a feature descriptor which is used for object detection. It is done by taking a set of positive and negative samples of the object we need to detect and train a linear support vector machine on your positive and negative samples. The next step is to apply hard negative mining. We will receive a set of false-positive samples and now re-train the classifier with these samples. Finally, the classifier will be trained. There are a lot of objects tracking algorithms in open CV like boosting tracker, Multiple Instance Learning (MIL) tracker, Tracking with Correlation Filters (KCF) tracker and so on. But among them the best tracker algorithm with more accuracy is CSRT tracker. That is the reason for using this tracking algorithm in this project. It is called as Discriminative Correlation Filter (with Channel and Spatial Reliability). It has higher object tracking accuracy and can tolerate slower Frames-Per-Second (FPS) throughput.

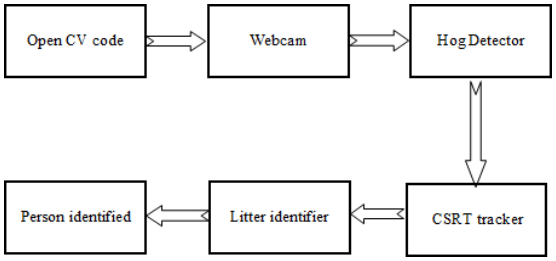

Figure 1. Block diagram for person face detection methodology.

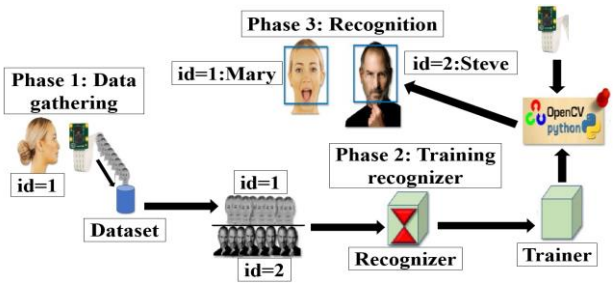

Figure 2. Person identification using face detection process.

Fig. 1 and Fig. 2 show the working flow of face detection and person identification.

IV. Implementation

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning program library. OpenCV was built to supply a common framework for computer vision applications and to quicken the use of machine discernment within the commercial items. Being a BSD-licensed item, OpenCV makes it simple for businesses to utilize and adjust the code. The library has more than 2500 developed algorithms, which include a detailed collection of computer vision and machine learning algorithms, both classic and state-of-the-art. Such algorithms can be used to detect and recognize faces, identify objects, classify human behavior in images, monitor camera movements, monitor objects moving, extract 3D object models, create stereo 3D point clouds.

NumPy could be a library for the Python programming dialect, including support for large, multi-dimensional clusters and frameworks, in conjunction with an expansive collection of high-level scientific capacities to function on these clusters. The precursor of NumPy, Numeric, was initially made by Jim Hugunin with commitments from a few other engineers. In 2005, Travis Oliphant made NumPy by joining highlights of the competing Numarray into Numeric, with broad adjustments.

Dlib may be a general reason cross-platform program library composed within the programming dialect C++. Its plan is intensely impacted by thoughts from plan by contract and component-based program designing. Hence it is, to begin with and first, a set of free computer program components. It is open-source program discharged beneath a Boost Program License.

Since improvement started in 2002, Dlib has developed to incorporate a wide assortment of devices.

As of 2016, it contains computer program components for managing with organizing, strings, graphical client interfacing, information structures, direct variable-based math, machine learning, picture preparing, information mining, Extensible Markup Language (XML) and content parsing, numerical optimization, Bayesian systems, and numerous other errands. In later a long time, much of the advancement has been centered on making a wide set of statistical machine learning devices and in 2009 Dlib was distributed within the Diary of Machine Learning Investigate [9]. Since at that point it has been utilized in a wide extend of domains.

First, we need to find the faces to be recognized. Secondly concentrate for each face to be able to identify that, it is still the same person even if a face is turned in different direction or in poor light. Third, you can find off the unique facial characteristics which you can use to tell for failure. Finally equate the unique characteristics of a person with all of the individuals you still know recognize to decide the name of the person.

Finding all the Faces-Face detection [10] is the basic step in our process. Face discovery could be an awesome highlight for cameras. When the camera can automatically choose out images, it can make beyond any doubt that all the images were in center some time recently it takes the images. But we'll utilize it for a diverse taskrecognizing of the picture we need in the following step in this process. Face location went standard within the early 2000's when Paul Viola and Michael Jones formulate to distinguish images that was quick sufficient to work on all cameras. Be that as it may, much more solid arrangements exist presently.

To discover faces in an image, we start by turning our picture dim since we don't require color information to discover face image. Fig. 3 is used as sample image.

Figure 3. Sample image for face detection.

As we realize that OpenCV loads a picture in PGR format to recognize its true color, we need to translate in to RGR format [11]. For all of this we need below function. We might be utilizing the detect Multi scale module of the classifier. This work will return a rectangle with coordinates(x,y,w,h) around the recognized confront. The work has two imperative parameters which ought to be tuned concurring to the information. Scale factor. There might be some faces in a group picture that are close to the camera than others, these faces will of course more noticeable than that of the ones behind. Min-Neighbors: It specifies to rectangle about having number of neighbors of a face.

Our next step is to use open CV our second step is to loop over all the dimensions returned and construct rectangles around them. Hence, we will draw a green rectangle with a thickness of 2. We conclude that the original colored pictured has been detected correctly or not. Presently we are at the crucial stage of the problem. Where we are dealing with the encoding faces. The easiest way to confront acknowledgment is to straightforwardly contrast the obscure confront we will be finding individual images in the previous step which are labeled. When we discover an already labeled confront which can be viewed as exceptionally comparative to our obscure confront, it should be similar individual. There's really a colossal issue with this way of approach. A location such as any social media with a large number of clients and an enormous data of photographs can't conceivably circle through each previously pointed confront which compares it to each recently transferred image.

The process of recognizing faces can be done in seconds, not hours. What we require may be a way to extricate some essential estimations from each confront. At that point we may degree our obscure confronts the same way and discover the known con- front with the closest estimations. This process of training works by seeing into three face pictures overall at once: Upload a trained face picture of a familiar person. Upload another image of the same person. Upload an image of another person who differs from above image (Fig. 3).

In the entire process, it is the simplest step. The task we need to perform in the step is to recognize the person in the data of that is identity of people who are closer to our test picture. Basically, we can give the input in two ways one is through web camera and other one is through video which will be uploaded. Here we will be taking the picture of the person to be identified and then we will be recognizing the name of person. The output will be the name of the person as in Fig. 4.

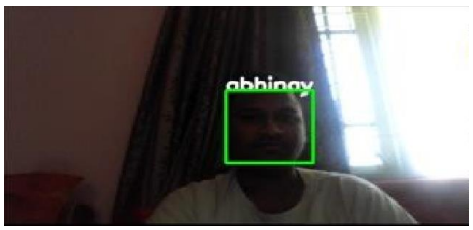

Figure 4. Face detection along with name.

V. Results and Discussion

Face recognition usually involves a process of three steps generally, Object detection is one of a kind in computer technologies, which is connected to the image processing & computer vision, and it interacts with detecting instances of an object such as human faces, building, tree, car, etc. The primary aim of face detection concept or algorithm is to determine whether there is any face in an image (or) live video or not.

Open CV has its own face detection neural network with high accuracy, which is also used to extract images

of people's face from number of images, or a live video which can also be called as cropping the faces of people in an image or video [12] for more precise and accurate details. Face recognition first involves location of a face, cropping it and compares with the faces in the database that is already saved, finally gives the result. Sometimes if the people in the video are large then it may not be able to identify it, then a kind of multi-face [13] challenging dataset is required. Fig. 5 gives object tracking [14].

We clearly see the object identifying presented in the hand is in Fig. 5. The object is indicated in the rectangular green box. Here, only the object is identified and person carrying it is not shown. The accuracy score observed is 12.17.

Figure 5. Person detection along with object using CSRT tracker.

We can see the detection of object along with face detection from Fig. 5. Both person and object are validated at a time. Sample dataset is taken from Kaggle. We defined the person along with object detection and their accuracy score as 12.7 in Fig. 5. To generate this experimental result, we have designed an Open CV with a Python. Here, the result is generated with Person name as 'Shivani', and accuracy score as 68.69%.

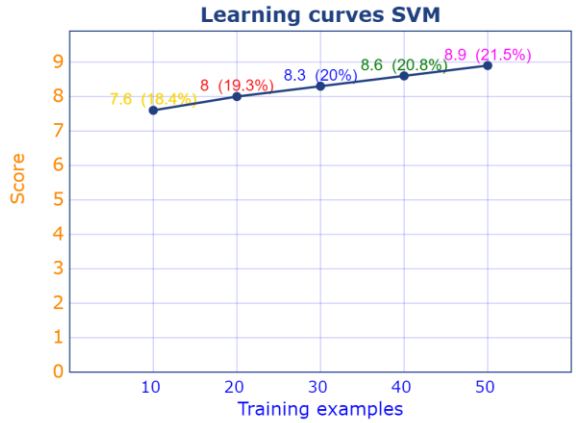

Figure 6. Training model graph generating accuracy for SVM model and training score.

It is assumed that the graph is constantly increasing and can infer that with the increase of training samples the score or the performance is increasing shown in Fig. 6. The cross-validation [15] score is also increasing so it is clear that the model is not over-fitting. We have defined the Learning curves using SVM model in Fig. 7 and also defined the Scalability and Performance of the model in Fig. 8 and Fig. 9. The result also shows the Training score and Cross Validation score Linear Curves SVM as 8.9.

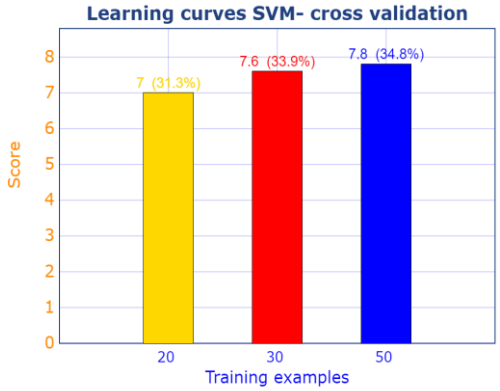

Figure 7. Training model graphs generating accuracy for SVM model, training score and cross validation.

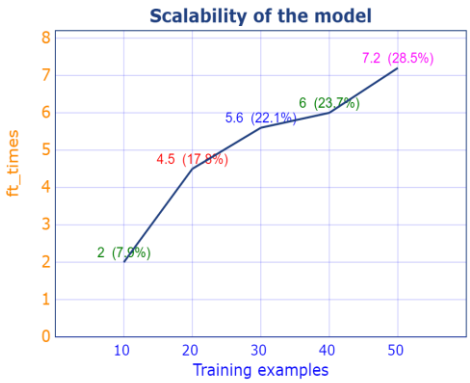

Figure 8. Training model graphs generating the scalability of the model.

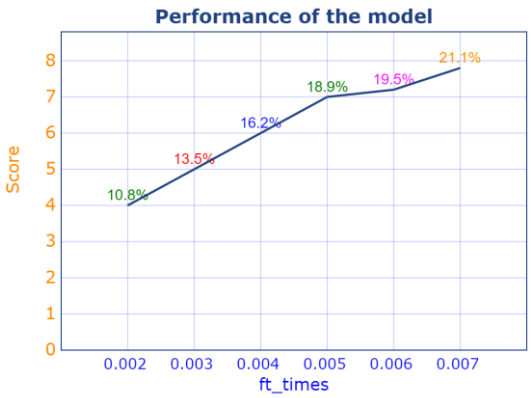

Figure 9. Training model graphs generating the performance.

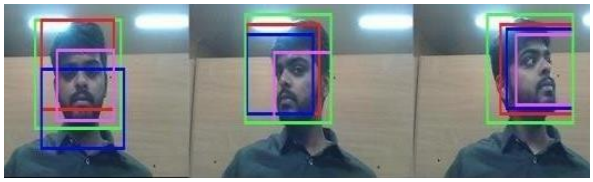

Figure 10. Image detection tracking on low light.

We have defined the lighting and scaling of the image presence in Fig. 10. The object is identified in the low light in different angles like Face looking forward, Face looking left side, and Face looking right side. Here, three types of boxes are used to define features of detected object like hair, eyes, nose, mouth, ears and fore head etc under in different lighting conditions. The features detected in colored boxes like green box-full object detection, pink box-detecting eyes, nose and mouth, red box-to detect ears nose, mouth and eyes etc.

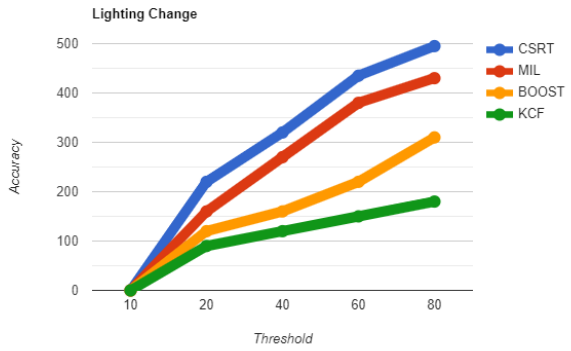

Figure 11. Algorithm accuracy comparison of CSRT, MIL, BOOST, KCF.

We have generated the accuracy scores comparison with CSRT, MIL, BOOST and KCF models shown in Fig 11. The highest accuracy score is achieved by CSRT shown in the graph with accuracy. The result is generated to test the lighting condition.

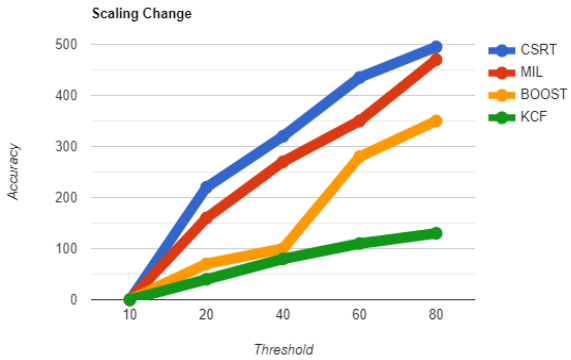

Figure 12. Scaling change accuracy graph.

Fig. 12 shows the tracking of object by different trackers in distant scaling, where the trackers having a tough time tracking the object (CSRT seems to be consistent over other trackers). This figures also shows the average of both lighting and scaling (Bad lighting condition and a bit more distant), CSRT seems to be fast and tops out the average accuracy.

We generated the Scaling change accuracy graph for the CSRT, MIL, BOOST and KCF models are sketched in Fig. 13. The CSRT has the maximum scaling change affect when compared with remaining models. The average accuracy is compared between the CSRT, MIL, BOOST and KCF, and CSRT having the highest value of 0.99 (99%) is mentioned in Fig. 13. The average scores of each model are compared in all aspects of object detection.

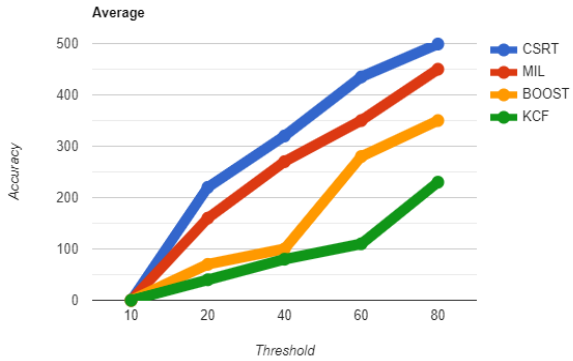

Figure 13. Average accuracy graph.

VI. Conclusion

Using proposed method, we improve the face recognition system under illumination variation and non-frontal view. Proposed approach is very simple in term of calculation. Improve Speed of recognition. It required only one scanning without any need to a complicated analysis. To move forward the comes about advance able to utilize more profound learning approach, for that we require a parcel of learning data, so these gotten to be the longer-term scope. Human action acknowledgment isn't so distant from being fathomed, but with the advances utilizing presently like profound learning and Convolutional Neural Systems, it can be possible.

This application can also be used for surveillance purpose in home and offices by making certain changes in the code. Smart attendance system is popular and developing aspect nowadays, this can be of certain help to implement the feature or system. A penalty payment procedure can also be introduced to punish the offenders. Same application can be applied to mass littering with a powerful engine with huge number of resources.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All the authors in this paper contributed their efforts for implementation of the paper. Srikanth Bethu has done proposed work implementation. M. Neelakantappa, A. Swamy Goud, B. Harikrishna and P. N. V. Syamala Rao M. implemented results part.

References

[1] J. Shelton and G. Dozier, "Genetic based LBP feature extraction and selection for facial recognition," in *Proc. the 49th Annual Southeast Regional Conference*, 2011.

[2] R. Rouhi, M. Amiri, and B. Irannejad, "Review on feature extraction techniques in face recognition," *Signal & Image Processing an International Journal*, vol. 3, no. 6, pp. 1–14, 2012.

[3] Y. Lu, J. Zhou, and S. Yu, "A survey of face detection, extraction and recognition," *Computing and Informatics*, vol. 22, no. 2, pp. 163–195, 2003.
[4] R. D. Agushinta, A, Suhendra, S. Madend, and H. S. Suryadi, "Face component extraction using segmentation method on face recognition system," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 2, 2010.
[5] D. Maturana, D. Mery, and A. Soto, "Face recognition with LBP, spatial pyramid histograms and naïve bayes nearest neighbor classification," in *Proc. 2009 International Conference of the Chilean Computer Science Society*, 2011.
[6] P. Sharma, K. V. Arya, and R. N. Yadav, "Efficient FR using wavelet based generalized neural network," *Signal Processing*, vol. 93, issue 6, pp. 1557–1565, 2013.
[7] M. Khan, S. Chakraborty, R. Astya, and S. Khepra, "Face detection and recognition using open CV," in *Proc. International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2019.
[8] S. S. Charan and G. Saini, "Pedestrian detection system with a clear approach on Raspberry pi 3," in *Proc. International Conference on Inventive Research in Computing Applications*, 2018.
[9] X. Farhodov, O.-H. Kwon, K. W. Kang, S.-H. Lee, and K.-R. Kwon, "Faster RCNN detection based open CV CSRT tracker using drone data," in *Proc. 2019 International Conference on Information Science and Communications Technologies (ICISCT)*, 2019.
[10] K. Goyal, K. Agarwal, and R. Kumar, "Face detection and tracking using open CV," in *Proc. International Conference on Electronics, Communication and Aerospace Technology*, 2017.
[11] A. W. Senior, "Recognizing faces in broadcast video," in *Proc. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999.
[12] S. R. Dubey and S. Mukherjee, "A multi face challenging dataset for robust face recognition," in *Proc. 2018 15th International Conference on Control*, 2018.
[13] D. T. P. Hapsari, C. G. Berliana, P. Winda, and A. Soeleman, "Face detection using haar cascade in difference illumination," in *Proc. 2018 International Seminar on Application for Technology of Information and Communication*, 2018.
[14] I. Paliy, "Face detection using Haar-like features cascade and convolutional neural network," in *Proc. 2008 International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science*, 2008.
[15] A. Dehghani, D. Moloney, and I. Griffin, "Object recognition speed improvement using BITMAP-HoG," in *Proc. 2016 IEEE International Conference on Image Processing (ICIP)*, 2016.