International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 2002-2008 www.ijaem.net ISSN: 2395-5252

Face, Expression And Gesture Recognition and Compilation in Database Using Machine Learning

Prof. Prashant Wakhare[1], Vaishnavi More[2], Rutuja Surdi[2],Vishwadip Ingale[2], Kajal Patil[2]

[1]*Professor at Information Technology Department, AISSMS Institute Of Information Technology, Pune,Maharashtra,India*

[2]*B.E Scholar, Information Technology Department, AISSMS Institute Of Information Technology, Pune,Maharashtra,India*

Submitted: 01-07-2021 Revised: 13-07-2021 Accepted: 16-07-2021

ABSTRACT: In today's scenario, numbers of crimes have increased day by day. At many public places government has placed many CCTV cameras so police can get that CCTV footage to identify the suspects but sometimes it becomes difficult to recognize the criminals So here we have come up with a solution to make this process smooth, easier than the traditional one. The system which automates all the suspect recognition process and provides better solutions to reduce the increasing rate of crimes. We plan to design a system to capture face, expressions and gestures of the targeted people (Criminals) through distributed CCTV System and are maintaining it in a database along with time and location stamp. The compiled database will be used to identify suspects from video clips of crime related CCTV footage captured series of CCTV Systems located on routes and close to scene of crime. This research discusses the various types of methodologies that can be used to identify the suspects which are captured in CCTV footage and convert it into useful information for further analysis of particular crime cases.

KEYWORDS: Artificial Intelligence, MachineLearning,OpenCV,CCTV,OpenFace,LSTM,YOLOV3

I. INTRODUCTION

Security is always a main concern in every domain, due to a rise in crime rate in a crowded event or suspicious lonely areas. Abnormal detection and monitoring have major applications of computer vision to tackle various problems. Due to growing demand in the protection of safety, security and personal properties, needs and deployment of video surveillance systems can recognize and interpret the scene and anomaly events play a vital role in intelligence monitoring. At many public places government has placed many CCTV cameras so Police can get that CCTV footage to identify the suspects but sometimes it becomes difficult to recognize the criminals So here we have come up with a solution to make this process smooth and easier than the traditional one.We plan to design a system to capture face, expressions and gestures of the targeted people (Criminals) through distributed CCTV System and are maintaining it in a database along with time and location stamp. The database so compiled will be used to identify suspects from video clips of crime related CCTV footage captured series of CCTV Systems located on routes and close to scene of crime.Video surveillance systems using Closed Circuit Television (CCTV) cameras, is one of the fastest growing areas in the field of security technologies. However, the existing video surveillance systems are still not at a stage where they can be used for crime prevention. The systems rely heavily on human observers and are therefore limited by factors such as fatigue and monitoring capabilities over long periods of time. This work attempts to address these problems by proposing an automatic suspicious behaviour detection which utilises contextual information.

II. COMPARISON AND ANALYSIS

According to survey, all the techniques which are used by previous authors have some limitations. To deploy this application the main challenge we were facing is to combine all this separate modules into one. So that we will get all models altogether and the efficiency will be increased accordingly. The second challenge is that we are using supervised machine learning algorithms so we need large amount of data for comparison and processing of each individual model. So we have found some algorithms which uses less amount of data but acquires high accuracy as compared to other techniques which we have seen briefly in literature survey. Basically proposed system is divided into five modules they are respectively:

1.Face,expression detection
2.Weapon detection
3.Hand gesture recognition
4.Pose estimation.

All these modules are combined together to get the result in pie chart form for further analysis. And extension of this proposed system include Application Programming Interface which can be available publically for commercial and private users who need it so that can be used to deploy this kind of application that needs criminal's data for identification. The existing system does not have this feature. We tried to find out which systems are mostly preferred for analysing the details of the selected papers. However, we found that most of the papers containing general definitions and there were insufficient information on the technical implementation details. So we figured out following algorithms that can be used to reach higher accuracy and better results.

For Face and expression recognition we have decided to use OpenFace algorithm which is a Python and Torch implementation of face recognition with deep neural networks and is based on the CVPR 2015 paperFaceNet: A Unified Embedding for Face Recognition and Clustering by Florian Schroff, Dmitry Kalenichenko, and James Philbin at Google. Torch allows the network to be executed on a CPU or with CUDA.

For Weapon Detection we have planned to use YOLOv3 (You Only Look Once, Version 3) which is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. Versions 1-3 of YOLO were created by Joseph Redmon and Ali Farhadi. The first version of YOLO was created in 2016, and version 3, was made two years later in 2018. YOLO is implemented using the Keras or OpenCV deep learning libraries.For hand gesture recognition we have planned to use OpenCV python libraries this can be implemented by following sequence of methods for recognizing hand gestures. Algorithm includes hand detection, fingers and palm segmentation, fingers recognition and finally whole hand gesture recognition.For Pose estimation module in proposed system we have planned to use LSTM algorithms: Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more.LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field. So this way all these models will be combined together to get all these functionalities in one application.

III. EXPERIMENTAL RESULTS

1.FACE DETECTION(OpenFace Algorithm):

The following overview shows the workflow for a single input image of Sylvestor Stallone from the publicly available LFW dataset.

1. Detect faces with a pre-trained models fromdlibor OpenCV.
2. Transform the face for the neural network. This repository uses dlib's real-time pose estimation with OpenCV's affine transformation to try to make the eyes and bottom lip appear in the same location on each image.
3. Use a deep neural network to represent (or embed) the face on a 128-dimensional unit hypersphere. The embedding is a generic representation for anybody's face. Unlike other face representations, this embedding has the nice property that a larger distance between two face embeddings means that the faces are likely not of the same person. This property makes clustering, similarity detection, and classification tasks easier than other face recognition techniques where the Euclidean distance between features is not meaningful.
4. Apply your favorite clustering or classification techniques to the features to complete your recognition task. See below for our examples for classification and similarity detection, including an online web demo.

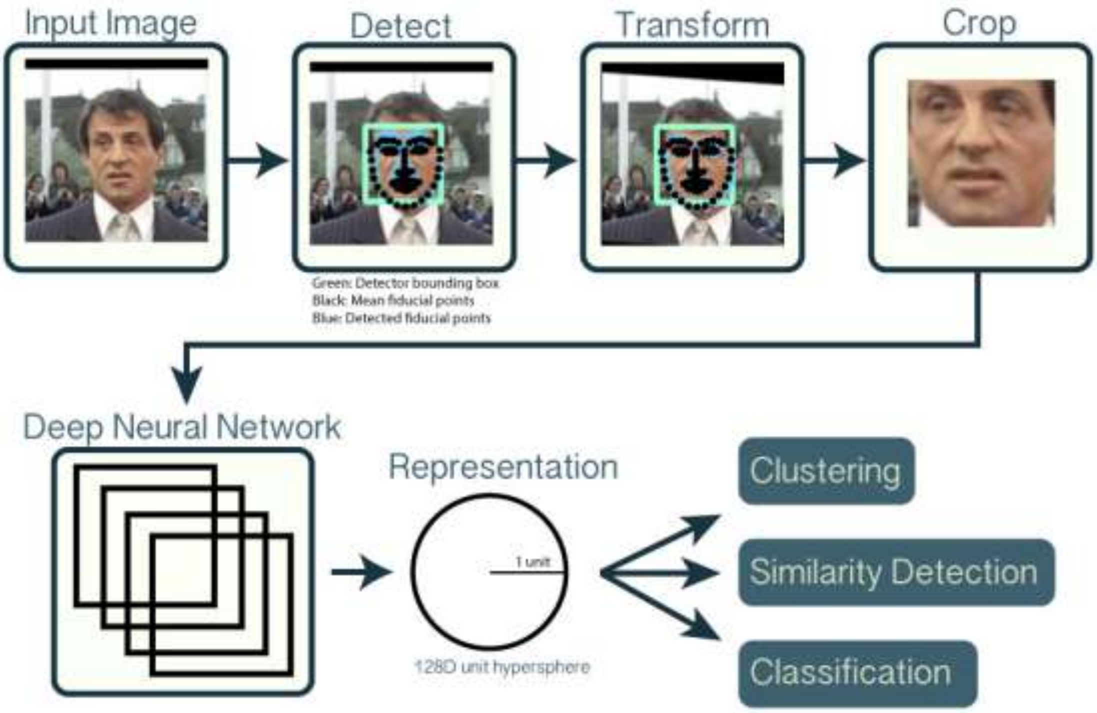


Fig.1.1

RESULTS:

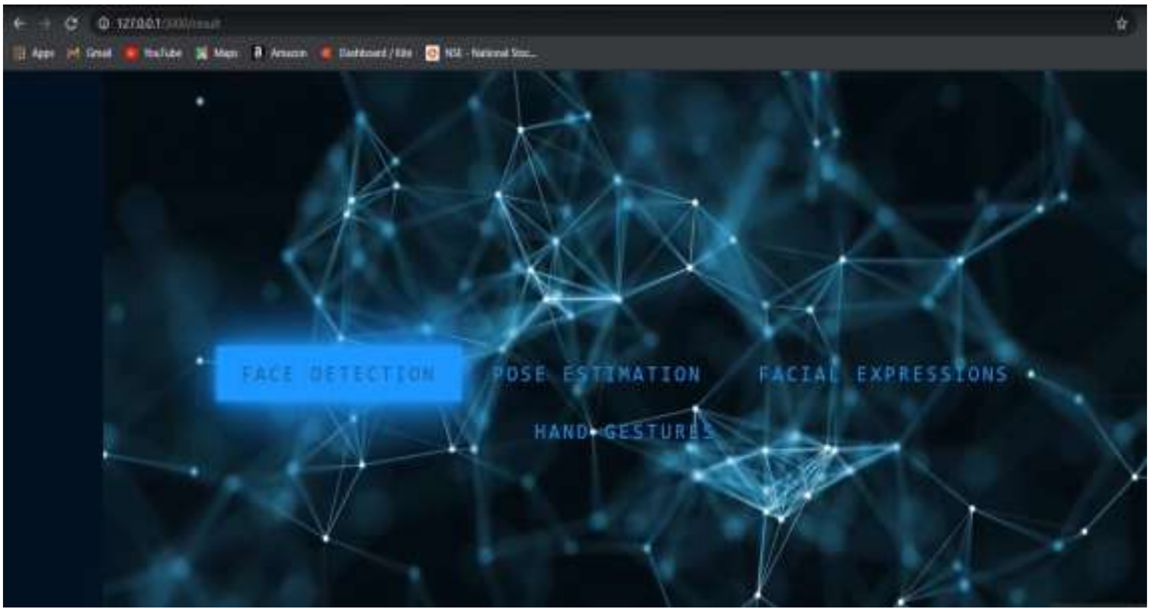


Fig.1.2

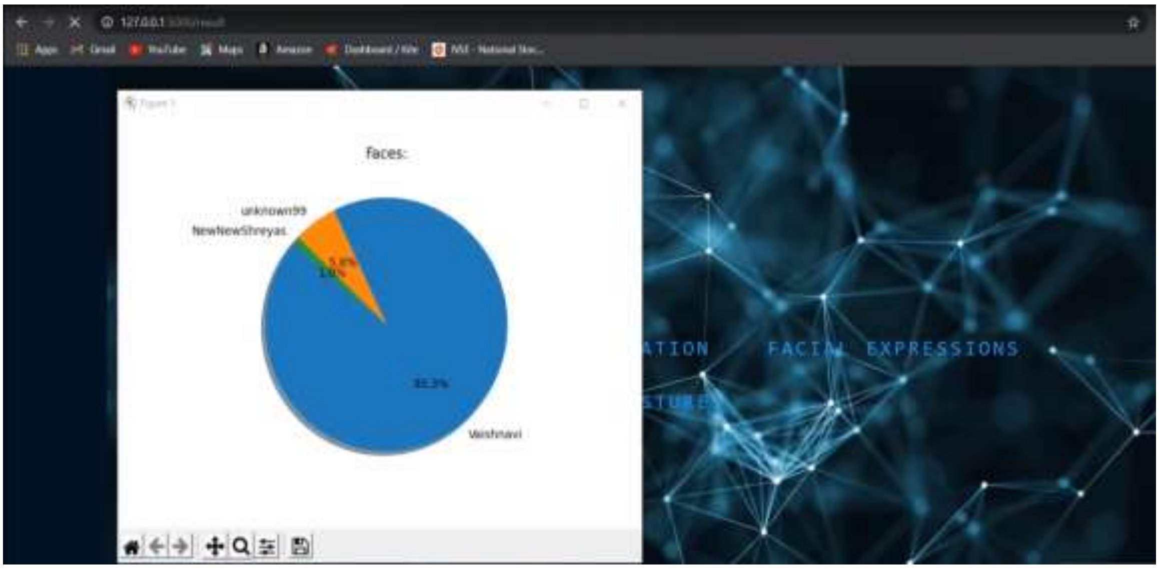


Fig.1.3

2.WEAPON DETECTION (YOLOV3 Algorithm):

YOLO is a Convolutional Neural Network (CNN)for doing object detection. CNNs are classifier-based systems that can process input images as structured arrays of data and identify patterns between them. YOLO has the advantage of being much faster than other networks and still maintains accuracy.It allows the model to look at the whole image at test time, so its predictions are informed by the global context in the image. YOLO and other convolutional neural network algorithms "score" regions based on their similarities to predefined classes.High-scoring regions are noted as positive detections of whatever class they most closely identify with. For example, in a live feed of traffic, YOLO can be used to detect different kinds of vehicles depending on which regions of the video score highly in comparison to predefined classes of vehicles.

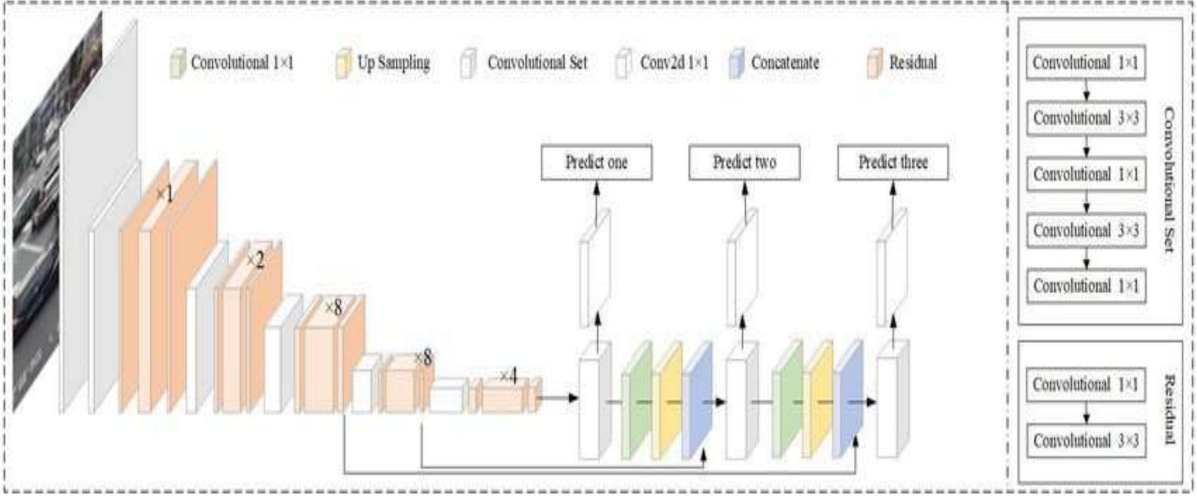
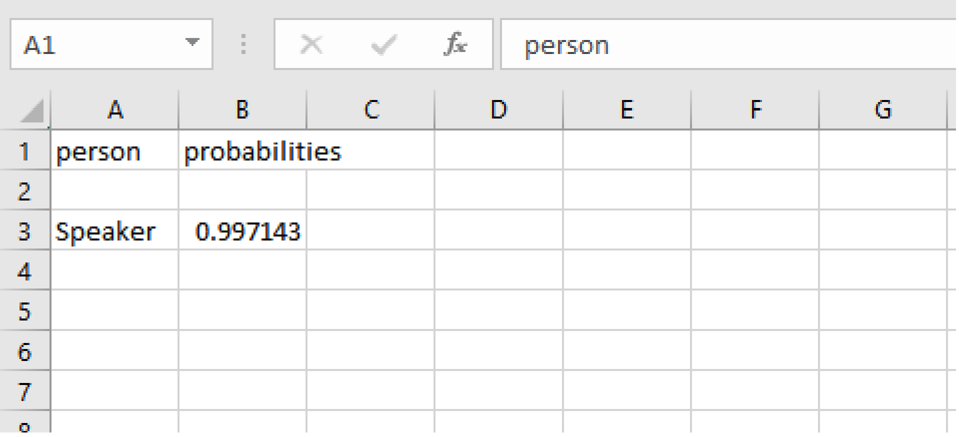


Fig.2.1

RESULTS:



	A	B	C	D	E	F	G
1	person	probabilities					
2							
3	Speaker	0.997143					
4							
5							
6							
7							

Fig.2.2

3.HAND GESTURE RECOGNITION(CNN Algorithm):

Convolution Neural Networks are part of deep learning technologies which is high level of machine understanding technique. So the neural networks are layered processing algorithm where it contains input layer, output layer and several middle layers. So this middle layer comprises of processing layers like convolution, pooling, recurrent, dropout, noise, normalization and many more. So in the model which we are going to implement consist of some of these layers that would be processing images faster and loading the features based on the number of times the training is going on / number of samples that it’s been training with. Keras is the neural network library that will be imported to make CNN work on our system.

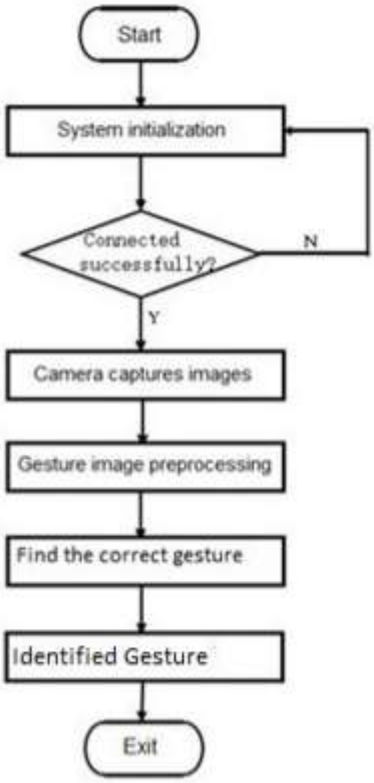


Fig 3.1

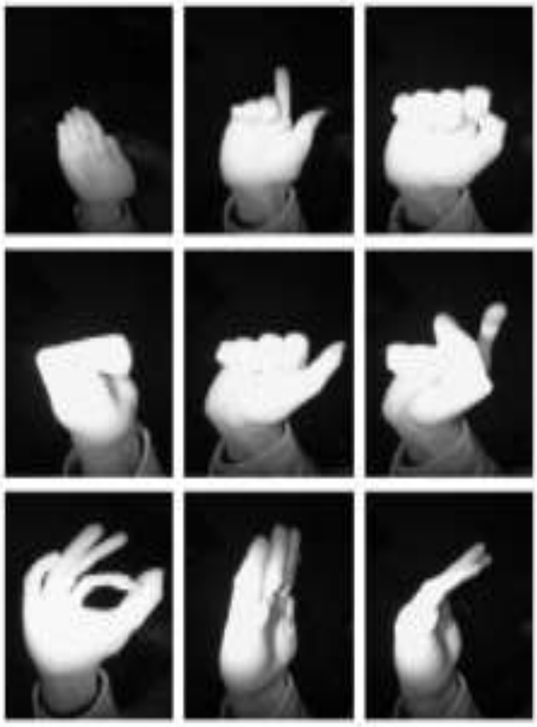


Fig 3.2

RESULTS:

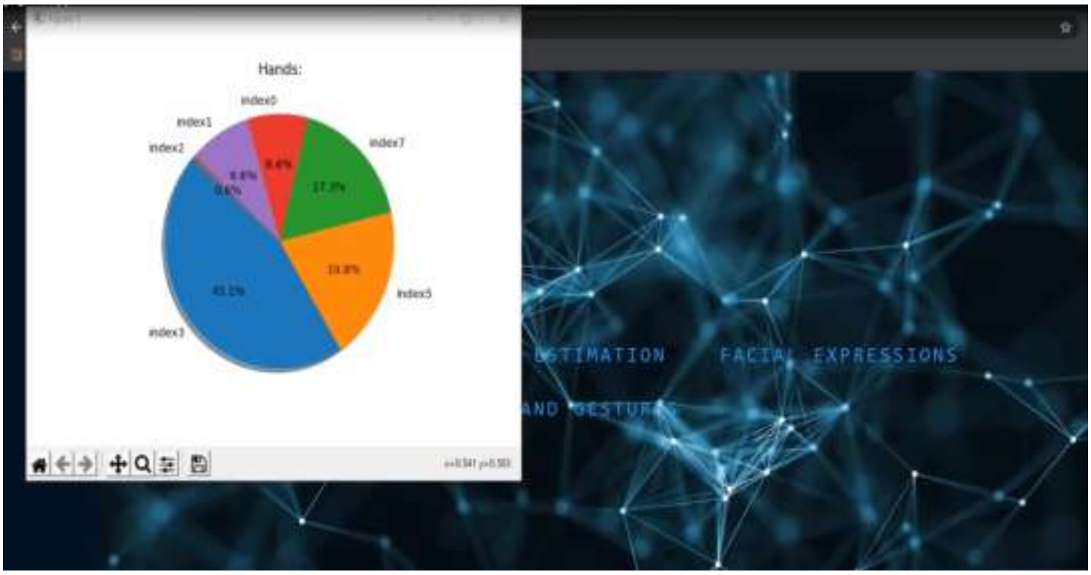


Fig.3.3

4.POSE ESTIMATION(LSTM Algorithm):

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems.This is a behavior required in complex problem domains like machine translation, speech recognition, and more.LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field.LSTMs on the other hand, make small modifications to the information by multiplications and additions. With LSTMs, the information flows through a mechanism known as cell states. This way, LSTMs can selectively remember or forget things. The information at a particular cell state has three different dependencies.

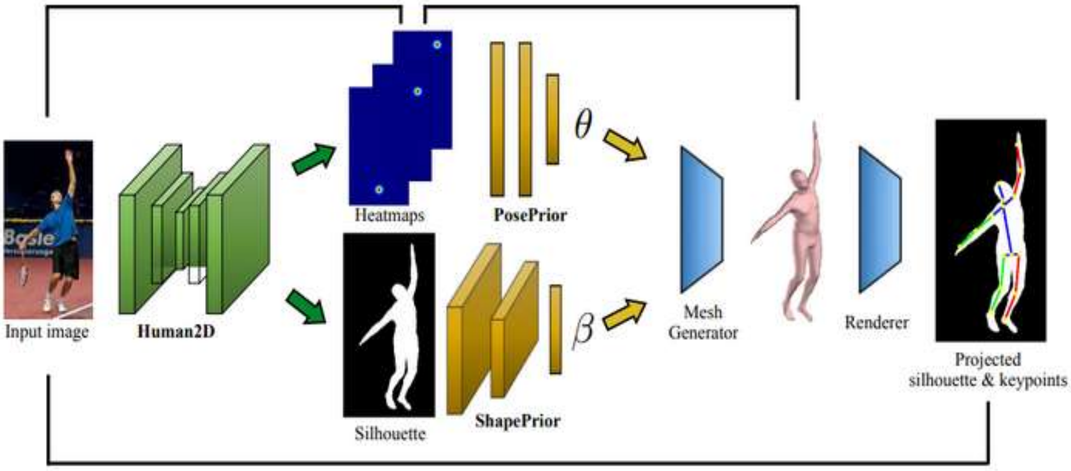


Fig.4.1

RESULTS:

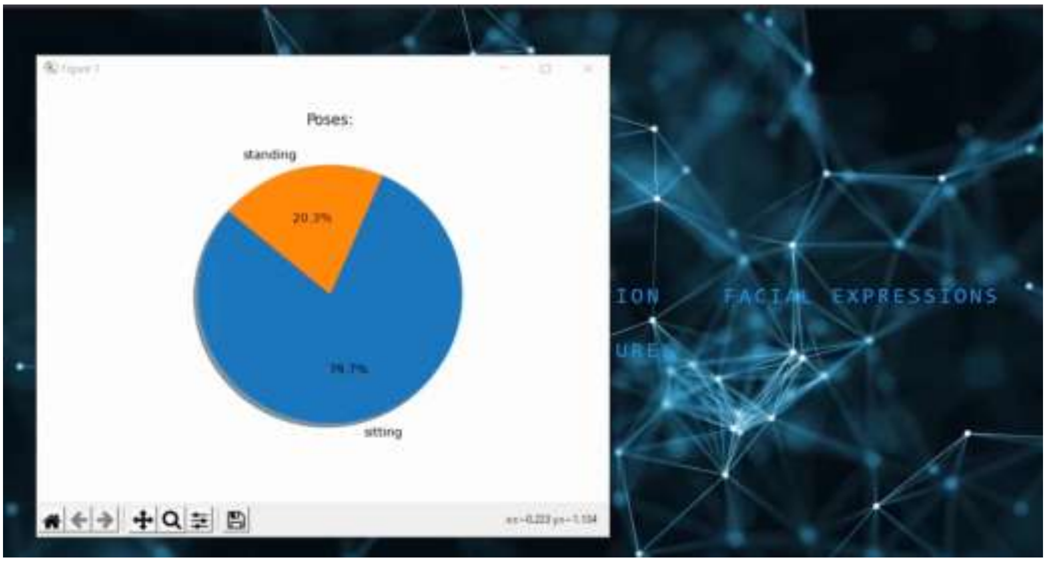


Fig.4.2

IV. CONCLUSION

As existing system contains all the separate moduleswhich user needs to run manually. This will make confusion to perform all this tasks and store the results accordingly.It is very time consuming because of its manual approach.And in this fast moving world where CCTV is only capturing the scenes, the system aims at using these datalike human face, expression, pose, gesture and what is human holding and converting it into usefulinformation using different machine learning algorithms. This information will help the authorities to gain a better insight inthe world of criminals/ suspects and get a readily available analysis.Thus it can act as both, crimedetection as well as a crime prevention system.Which help India to be crime free nation.

REFERENCES

[1]. M. Owayjan., R. Achkar. and M. Iskandar.: Face De-tection with Expression Recognition using Artificial NeuralNetworks. 3rd Middle East Conference on Bi-omedical En-gineering (MECBME), Beirut. IEEE MECBME, 115—119(2016).

[2]. J. Jayalekshmi and T. Mathew, Facial expression recog-nition and emotion classifi-cation system for sentiment analy-sis, 2017 International Conference on Networks & Advancesin Computational Technologies (NetACT), Thiruvanthapuram,India, 2017, pp. 1-8.

[3]. N.Paragios, (2018) Computer vision and understanding. Vol 116. pp 102-114.

[4]. Guojen Wen, Zhiwei Tong, et.al, (2009), Man machine interaction in machining center. International workshop on intelligent systems and applications. pp 1-4.

[5]. S.D. Bharkad, et.al. (2017). international conference on computing methodologies and communication, pp 1151-1155.

[6]. Nadhir Ben Halima and Osama Hosam. 2016. Bag of Words Based SurveillanceSystem Using Support Vector Machines. 10 (04 2016), 331–346.

[7]. D. M. Sheen, D. L. McMakin, and T. E. Hall. 2001. Three-dimensional millimeter-wave imaging for concealed weapon detection.IEEE Transactions on MicrowaveTheory and Techniques49, 9 (Sep 2001), 1581–1592. https://doi.org/10.1109/22.942570

[8]. S. Song and J. Xiao, "Tracking Revisited using RGBD Camera: Unified Benchmark and Baselines,"Princeton University Proceedings of 14th IEEE International Conference on Computer Vision (ICCV2013)

[9]. UCLAdatabase http://users.eecs.northwestern.edu/~jwa368/