



Vision

Mr. Shailendra Singh, Kartikeya Gaur, Muskan Rajput, Vedika Verma

Department of Computer Science & Engineering, Inderprastha Engineering College, Uttar Pradesh, India

Abstract: As the name VISION suggests the major aim of this project is to see and learn. The visually impaired people in the world face a lot of problems in day-to-day life. They need either a human or a stick to guide them through their different daily life tasks. They often get hit by objects because they're not able to see them coming toward them. They're not able to recognize people and objects without touching or hearing them. Entertainment is also a luxury for them since they can hardly operate any device. This research paper depicts a project that can solve all the above problems and perhaps more in long run. This project has applications all over the defense industry, security systems, autonomous cars, robot development, and many more fields.

I. INTRODUCTION

Our project utilizes the YOLO model for real-time object detection with the use of an embedded camera and then by using the Pyttsx library and speech recognition python package, the detected objects are notified to the user through the voice message.

When implemented using a device having a camera and microphone/ speaker, this is going to guide and indicate to the visually challenged people about their surroundings in a very efficient manner.

Not only the objects but our project alrecongizesses the faces of the known persons who are appearing in front of the visually challenged person.

Our project is also capable of performing face recognition and object detection tasks by using a portable or mobile camera device coconnectedo to our model.

Therefore, this prototype implementation will be used by the engineers for the welfare of theually impaired people or those who are unable to see properly due to various vision disorders.

1.1 PROBLEM DEFINITION

This project focuses on making a device that can be used to detect faces and objects through a webcam, smartwatch, mobile phone, or any other portable camera device at a very efficient speed to capture clear images. It not only detects the faces, and objectsand but also trains the dataset to identify if the same face or object is encountered again the person will be notified through a voice message. This project has applications all over the defense industry, security systems, autonomous cars, robot development, and many more fields.

1.2 BACKGROUND ABOUT PROJECT IDEA

We were traveling in the metro and saw an old blind lady passing by us. She was struggling to get inside the metro, and we felt so sad by seeing her condition so, we got an idea popped in our mind that we can do something for these kinds of people. So, we decided to make our college project for the same. The main reason for making this project was that this can be implemented in real-life scenarios and can help the needed ones.

1.3 FEASIBILITY STUDY, NEED, AND SIGNIFICANCE

The feasibility study is a major factor that contributes to the analysis and development of the system. The decision of the system analyst whether to design a particular system or not depends on its feasibility study. A feasibility study is undertaken whenever a possibility of the probability of improving IJCSN - International Journal of Computer Science and the existing system or designing a new system. A feasibility study helps to meet user requirements. a. Financial Feasibility: The goal of this product is clear and it is possible to make this project with all the required content keeping the project financially strict will be our main goal so that it could be implanted further on a large scale. b. Technical Feasibility: As far as machine learning is concerned a high processing system is required for maximum accuracy. Our goal is to achieve maximum throughput in less power consumption c. Operational Feasibility: The basic operation of what Image Processing is and how will it help in digit recognition should be reflected through this feasibility report. Financial Feasibility: The goal of this product is clear and it is possible to make this project with all the required content keeping the project financially strict will be our main goal so that it could be implanted further on a large scale.



i Technical-Feasibility: As far as machine learning is concerned a high processing system is required for the max. Accuracy. Our goal is to achieve maximum throughput in less power consumption. c. Operational-Feasibility: The basic operation of what Image Processing is and how will it help in digit recognition should be reflected through this feasibility report

SCOPE:

- The VISION plans on being the go-to tool for all visually disabled people so they can experience the best of the world around them without feeling left out or deprived of any senses. This tool will recognize people and objects around the person and inform him about those through voice assistance installed in the device.

1.4 NOVELTY OF PROJECT

Most of the pre-existing “Recognition” applications in the market do not use AI / ML in their functioning. Here are some of those applications:

- ✦ Be my eyes (2M+ downloads) It’s an app with millions of downloads but AI doesn’t have much role in it the volunteer-based app is used on a large scale. It uses a rear camera and a volunteer from the backend will tell whatever is facing the phone
- ✦ Seeing AI (IOS) / Google lookout (android) (100k + downloads) Can detect objects and text from images and live cam. Y BARD / AUDIBLE (10k + downloads) Book narrator, can read books for blind people.
- ✦ Cash Reader (100k + downloads) Can read cash and tell details about a specific currency (Free version only available in the US though).

Note - The development of these apps requires a group of well-trained Engineers, Volunteers, and plenty of time. The main ques is what does Vision do differently from these apps?

Here is the answer:

♣ Vision has 2 different modules:

1. Object detection
2. Face detection

None of the previously mentioned apps have all these features.

- ✦ Be MY EYES do work with real human assistance but our project does it with AI, which is completely different.
- ✦ Google lookout won’t tell who (person) is in front of the person but Vision does with its Face Recognition feature. BARD and AUDIBLE only narrate the books whereas Vision has its virtual assistant for the work.

II. LITERATURE REVIEW

Every object class has its special features that help in classifying the class. For example, all circles are round. Object class detection uses these special features. For example: When looking for circles, objects that are at a particular distance from a point (i.e. the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed.

A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin color and distance between eyes can be found.

The earliest pioneers of facial recognition were **Woody Bledsoe Helen Chan Wolf** and **Charles Bisson**. In 1964 and 1965, Bledsoe, along with Wolf and Bisson began work using computers to recognize the human face.

This project originated from an unnamed intelligence agency, therefore much of their work was never published. But it was later revealed that their initial work involved the manual marking of various “landmarks” on the face such as eye centers, mouth, etc.

These earliest steps into Facial Recognition by Bledsoe, Wolf, and Bisson were severely hampered by the technology of the era, but it remains an important first step in proving that Facial Recognition was a viable biometric.

Detection methods are neural network-based onon–neural approaches. For nonon–neural approaches, it becomes necessary to first define features using one of the methods below, then using a technique such as a Support Vector Machine (SVM) to do the classification.

On the other hand, neural techniques can do end–to–end object detection without specifically defining features, and are typically based on Convolutional Neural Network (CNN).

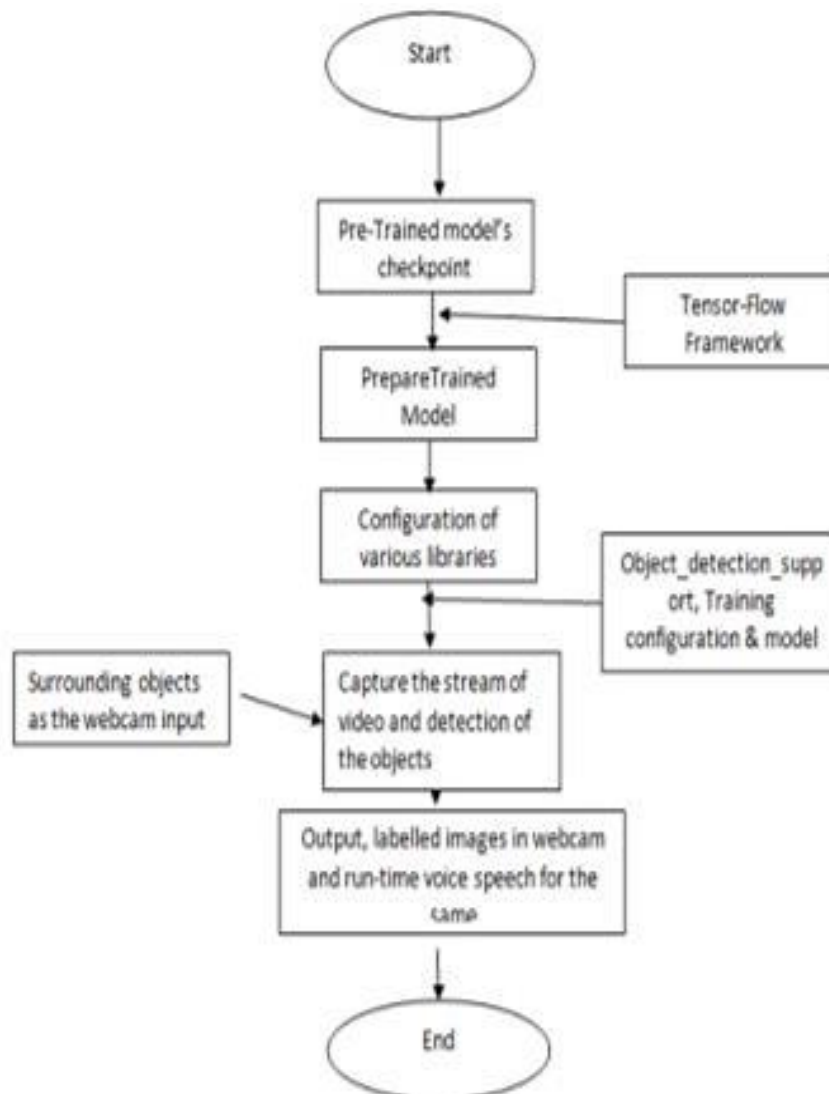
- Non-neural approaches: o Viola-Jones object detection framework based on Haar features. o scale-invariant feature



transforms (SIFT). o Histogram of oriented gradients (HOG) features.

- Neural network approaches: o Region Proposals (R – CNN, Fast R – CNN, Faster R – CNN, Cascade R - CNN). o Single Shot MultiBox Detector (SSD). o You Only Look Once (YOLO) o Single–Shot Refinement Neural Network for Object Detection (RefineDet) o Retina - Net o Deformable convolutional networks.

III. FLOW CHART



IV . METHODOLOGY

✦ **Yolo Model** was proposed by Joseph Redmond et al. in 2015. It was proposed to deal with the problems faced by the object recognition models at that time, Fast R-CNN is one of the state-of-the-art models at that time but it has its challenges such as this network cannot be used in real-time, because it takes 2-3 seconds to predicts an image and therefore cannot be used in real-time.

✦ **Open CV** is the huge open-source library for user vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even the handwriting of a human. When it enters grated with various libraries, such as Numpy, python is capable of processing the OpenCV array structure for analysis. To identify imapatternsern and their various features we use vector space and perform mathematical operations on these features.

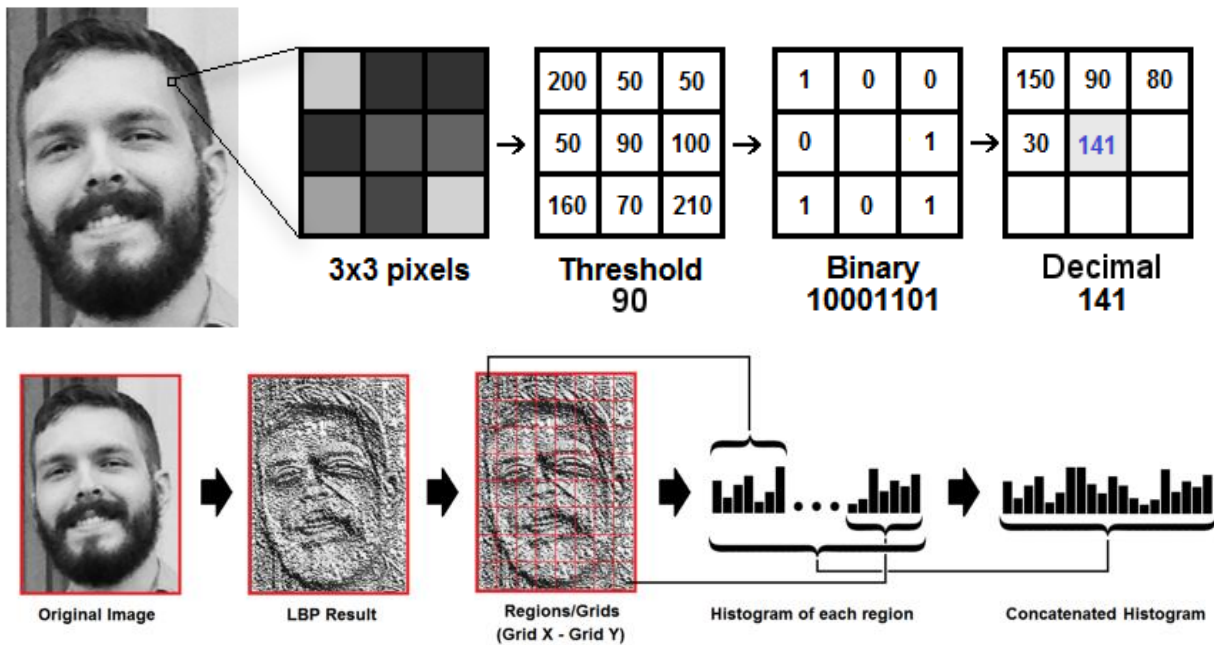


OpenCV Functionality:

- Image/video I/O, processing, display (core, imgproc, highgui)
- Object/feature detection (objdetect, features2d, nonfree)
- Geometry-based monocular or stereo computer vision (calib3d, stitching, videostab)
- Computational photography (photo, video, superres)
- Machine learning & clustering (ml, flann)
- CUDA acceleration (GPU)

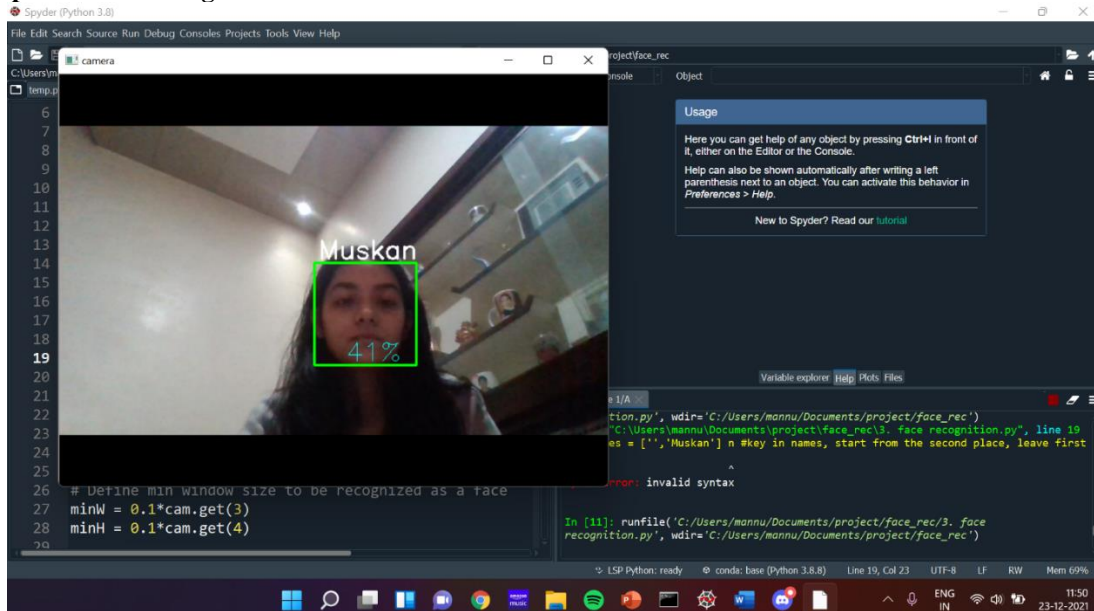
✦ **Local Binary Pattern (LBP)** is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.

✦ Using the LBP combined with histograms we can represent the face images with a simple data vector.

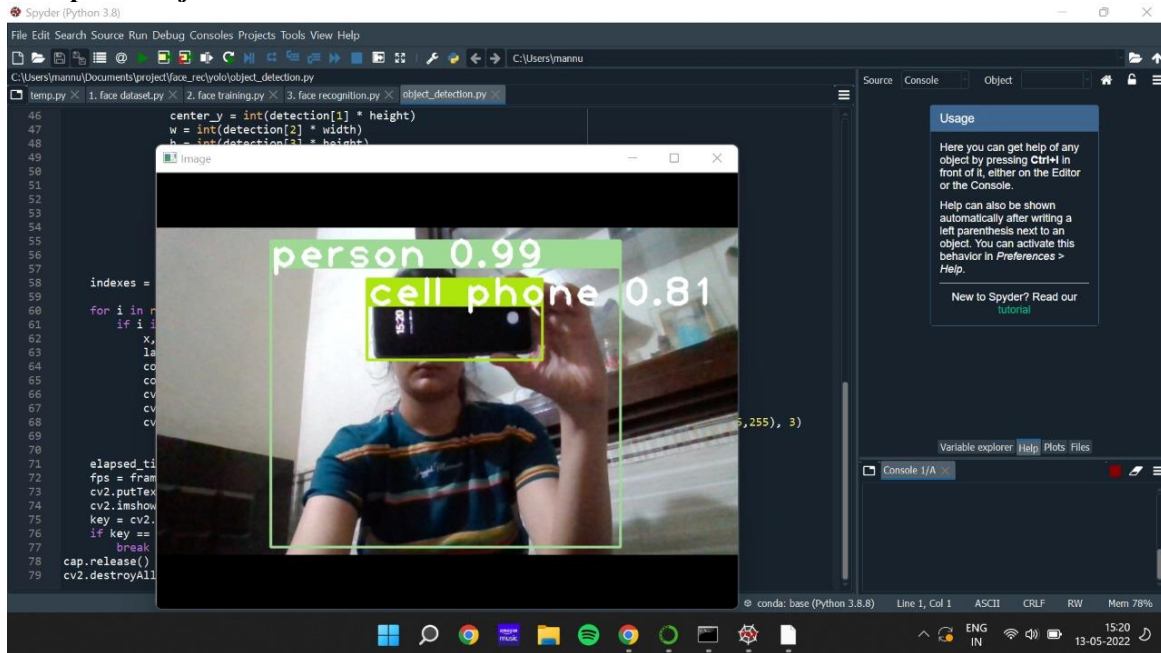


V. RESULT

7.1 Output of Face Recognition



7.2 Output of Object Detection



VI. CONCLUSION

- LBPH is one of the easiest face recognition algorithms.
- It can represent local features in the images.
- It is possible to get great results (mainly in a controlled environment).
- It is robust against monotonic grayscale transformations.
- It is provided by the [OpenCV](#) library (Open Source Computer Vision Library).
- To find faces in a real-time scenario we used OpenCV, and it also did a pretty amazing job. Its accuracy is around 80 % and it can detect faces in real-time.

With the combination of these 3, we can build a prototype that can be used by a blind person to get a virtual vision. With our efforts, we hope we can do something for the welfare of the visually impaired society.

VII. FUTURE WORK

In the future, we are planning to implement our work more realistically. We plan to make a fully working spectacle that has a camera, speaker, and microphone inbuilt so that the camera captures the image of the obstacle, and with the help of the speaker, the blind person can get information about who is in front of him. A Bluetooth device or some other connectivity tool will be used to send information to a phone. The phone will process the information with the help of our current model and then it will again send the outcome to the spectacle to narrate who is in front of the person.

VIII. REFERENCES

- <https://medium.com/@nikasa1889/the-modern-history-of-object-recognitioninfographic-aea18517c318>
- https://en.wikipedia.org/wiki/Object_detection
- <https://www.nec.co.nz/market-leadership/publications-media/a-brief-history-of-facialrecognition/>
- <https://ieeexplore.ieee.org/abstract/document/7961803>
- <https://www.sciencedirect.com/science/article/abs/pii/S0262885613001765>
- Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kurban, et al., "A benchmark and comparative study of video-based face recognition on cox face database", IEEE Transactions on Image Processing, vol. 24, pp. 5967-5981, Dec. 2015.
- 10. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016
- Google Scholar Research Papers
- Wikipedia