

# Smart Vision

Gopika K, Ani Sunny

*Abstract: Vision is the focal point to the world, translating the existence occasions. It says the manner by which see individuals and increase learning by perusing from regular scenes. On the off chance that scratch found on our glasses, it might appear to us that everybody around us has such a large number of scratches. The issue really lies with us, individuals who are visually impaired or outwardly disabled. They has no capacity to realize what are the things be occur around them and exceptionally rely upon others for everything. So a sentiment of weight is emerges. Smart vision is proposed here, which comprises of a few modules for helping the visually impaired individuals in an effective way. The structure includes Human face discovery framework and Text identification which make vision for outwardly tested. The current framework for supporting visually impaired individuals is utilize the gadget raspberry pi. Be that as it may, raspberry pi has little memory estimate. So it is troublesome for preparing and retraining. In my framework I tackled this issue by utilizing the idea quantization. Docker picture is utilized for retraining the entire picture, so it makes the framework in quicker way. The other real favorable position of the undertaking, utilization of constant gaming board UDDO x86 independent of gadget like raspberry pi and so on. The brilliant unit contains an eyeglass gave camera, headphone, receiver, battery and UDDO x86. The camera present at the eyeglass catch the proposed picture of the client and regular scenes as previews and exchange to the UDDOx86 where it gets handled and produce the ideal sound portrayals as yield utilizing headphone.*

*Index Terms: Face Recognition, Text Detection, Tesseract OCR, Eye aspect ratio*

## I. INTRODUCTION

There are more than 286 million individuals who are outwardly disabled and there are more than 40 million individuals who are thoroughly visually impaired. The visual capacities has constrained these people from totally seeing their prompt surroundings which has potential wellbeing concerns and furthermore brings down their personal satisfaction since they should use on a type of defensive guide to get around. At present, all together for outwardly impeded people to get around, they depend on white sticks, puppies and utilize individual human guides for help.

While these white sticks and guided puppies may enable the person to get around freely, they each have a typical downside. These gadgets have the absence of knowledge to give bearings to unvisited areas and can't totally caution people of obstruct questions in their region. A human gives this insight yet makes the outwardly disabled individual reliant on the human guide. India has world's biggest number of visually impaired individuals. Of the 38 million individuals over the globe who are visually impaired, more than 14 million are from India. Outwardly hindered individuals are hard to work in new conditions or unusual settings. These difficulties limit their autonomy and extend the hole among them and the regularly located populace. They are probably going to adjust their life to adjust to this specific condition, restricting their activities to counteract disappointments and dissatisfactions. In less controlled conditions, they may have issues in finding their way in obscure spots or in staying away from physical hindrances. In addition, it might be troublesome for them to pick garments and to purchase object likewise it is trying to discover a shop of a particular sort in a road they don't know well, or to enter a mail station, as they don't have a clue if the assistant is prepared to serve them. In India less development innovations or types of gear are available for visually impaired individuals, additionally less number of items propelled in India. Daze individuals utilize a white stick as an instrument for guiding them when they move or walk.

As of late there was no trend setting innovation to explore the visually impaired. They utilized generally material approaches to explore in space. The primary arrangements have been produced for the location of obstructions and risks in transit. Yet, around then likewise no gadget for face and content acknowledgment is created. In this manner, over the most recent couple of years there has been an enduring development of innovative work of strategies and frameworks to help outwardly disabled individuals. While there has been an impressive progression at the dimension of research discoveries, still most issues have been tended to freely. This lead to the improvement of gadgets or devoted applications that may, on a fundamental level, be viewed as intriguing guides, however it is for all intents and purposes difficult to envision a client receiving them in parallel.

**Revised Manuscript Received on December 22, 2018.**

**Gopika K**, Computer Science, Mar Athanasius College of Engineering, Kothamangalam, India

**Ani Sunny**, Computer Science, Mar Athanasius College of Engineering, Kothamangalam, India



Smart vision has a few modules for helping the visually impaired individuals proficiently. It gives a savvy and compact approach to help daze individuals. The framework enable outwardly weakened clients to just press a catch in film keypad, to switch between face acknowledgment and content acknowledgment mode and be guided there with the utilization of sound guidelines. Utilizing face acknowledgment innovation, the gadget will recognize cohorts, relatives and associates by giving a message on a headphone to tell the client continuously and has an office to label picture by name for new individuals. Messages in scenes and recordings contain critical and valuable data yet it not available by visually impaired. Utilizing content acknowledgment innovation, help the visually impaired individual for find explicit shops, cash, recognizing, distinguishing distinctive item name and their subtleties like cost and so forth.

## II. LITERATURE SURVEY

### A. Face Recognition

Face discovery is dependably a testing undertaking. In the Face recognition, the principal procedure is to include the picture then it is seek matches with the database. The input picture utilized in face acknowledgment process is additionally called as probe and the database utilized in face acknowledgment process is called as gallery. Match report is produced and the sub-populace is related to regard to grouping where new perceptions have a place. Face acknowledgment comprise of essentially three methodologies[1]. The neighborhood highlights like nose, eyes are portioned in feature based approach and it is considered as input information in face discovery for make face acknowledgment undertaking less demanding. The entire face taken as the contribution to the face recognition framework is holistic methodology. Crossover approach is gotten from the mix of highlight based and comprehensive methodology. In this methodology, entire face is considered for face discovery framework.

The discovery of face is happens first. The face discovery algorithms experience a few burns through different boxes, diverse element of appearances are considered. Facial highlights are recognized from the crate and score is relegated for giving a certainty level. On the off chance that face is understood, a layout is made, it dependent on a few factors, for example, the relative separation between the eyes and so forth. The created portrayal is contrasted and other recognized countenances[2]. The comparative proportions between separations on different purposes of the face yields a score on a logarithmic scale. The Close matches go is from 3 to 5, and esteem under 1 is ended up distinct non-matches. Facial acknowledgment innovation has a few components limit the effectiveness they are:

- Image quality: Quality of pictures is exceptionally rely upon working of facial acknowledgment calculations. The nature of video is very low contrasted and advanced camera. The top quality video is, best case scenario having 1080 dynamic output as a rule, it is 720p.
- Image estimate: The face-discovery calculation will focuses a face picture in an entire video. Face is

contrasted and the selected picture measure, it says the rate of acknowledgment of face. Little size of picture combined with an objective removed that is some specific separation from the camera.

- Face point: The face edge that focused is very contributes to acknowledgment score. An information is given to programming of acknowledgment utilizing numerous points of face are utilized. Other than frontal view, it influences layout for the face.
- Preprocessing: Always high definition video have low in resolution when compared with digital camera images, it occupies large significant amounts of disk space. As a result vast circle space is used. Colossal handling is done, so typically a small amount of 20 percent to 25 percent is really gone through an acknowledgment framework.

### B. Text Detection

Content discovery system can be grouped into two classifications principally picture investigation and machine learning based picture examination approaches. In the examination of pictures, discovery of content is significantly relies upon picture investigation. The basic stream fills in as edge recognition is performed on picture. The picture property highlights are as indicated by, spatial, and geometric relationships. It is reconstructed as potential content districts[3]. The districts of content are additionally broke down dependent on heuristic. In the second classification, the machine learning strategy is connected alongside picture examination systems. The machine learning fills in as the content pixel verifier. It is normally inputted after the examination of picture for lessening the bogus positive rate[4].

Text-recognition technology has certain feature limitations:

- Accuracy: The number of errors depends upon the quality and type of document, including the font used. Errors that occur include misreading letters, skipping over letters that are unreadable, or mixing together text from adjacent columns or image captions.
- Work-Arounds: The Existing Algorithm experiences issues separating between characters, for example, numerical zero and letter capital "O." Special text style can be utilized here.

## III. METHODOLOGY

### A. Components used for Hardware and Software

The procedure must be done both on programming and equipment, required hardware is as per the following:

- UDDO X86
- Camera
- Microphone
- Earphone
- Membrane keypad
- Battery



#### IV. PROPOSED SYSTEM

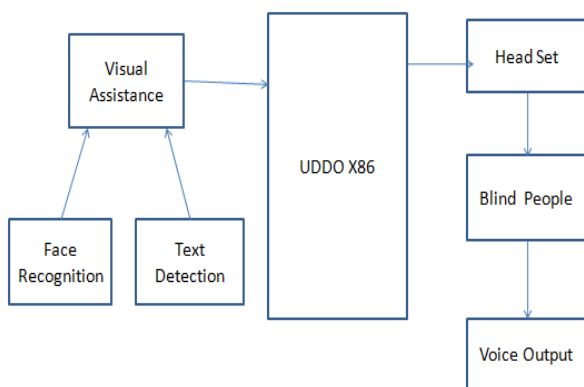


Figure 4.1: Block Diagram of the Proposed Method

Present a wearable device to assist visually impaired users in social interactions and gaining knowledge from natural scenes. The prototype is embedding a high resolution camera on specs and all the processing is done by UDDO X86 microprocessor.

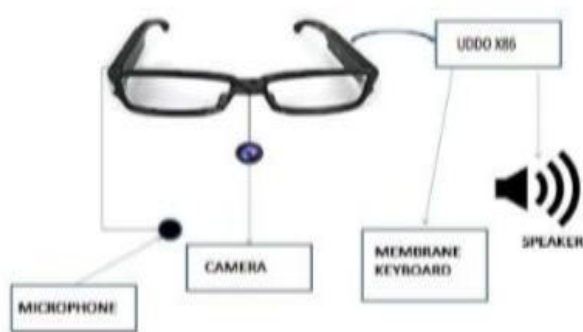


Figure 4.2: Proposed Architecture Diagram

##### A. Face Recognition

Using face recognition technology, the device will identify classmates, relatives and colleagues by giving a message on an earphone to notify the user in real time and has a facility to tag image by name for new people. If the same person appears in-front of the user then the name of a person is given as a message on earphone. If a new person came to interact with person then retrain the system for further communication process[5]. The device be developed is to support for recognize people easily and its societal benefit is to improve the access, integration and independence of the blind or visually impaired individuals in workspace or educational settings. Face recognition consists of mainly two steps detection and recognition[6]. In detection, consists of multiple steps:

1. Capture the faces from camera
  - a. Detecting faces within a predetermined middle boundary on screen having 200\*100 pixels dimension.

- b. In case of multiple faces within the determined boundary, the face closer to camera is focused on.
- c. If a focused face is in determined boundary, up to 5 frames are captured and multiplied to 60 frames per face is to reduce the overall face detection time.
- d. Frames that satisfy the following two conditions are saved.

i. Face Coverage  $\geq 40\%$

ii. Blurriness  $\geq 100$  Cycle per pixel(C/P)

- e. Eye Aspect Ratio is determined.

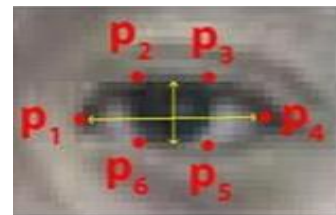


Figure 4.3: Eye aspect ratio

$$EAR = (||P2 - P6|| + ||P3 - P5||) / 2||P1 - P4|| \quad (1)$$

The EAR value with eyes remains constant and rapidly falls to zero when the eyes close between the blinks. This makes it an effective method for detecting live person or images present on advertisement board or others.

- f. Handling faces looking in different directions.

- i. Detect 68 specific points called landmarks that exist on every face such as the top of chin, outside edge of each eye, inner edge of each eyebrow etc.
- ii. Extraction of few basic measurements for each face such as size of each ear, the spacing between the eyes, the length of the nose, etc. It will reduce computation power and memory.
- iii. Verifying which person from the training dataset is the closest match with the test image. Linear SVM classifier is used because citing better prediction accuracy among all classifiers.

Recognition:

2 Finding face from the frame taken by the camera

- a. Input image is converted into form of grayscale.
- b. The grayscale image is further processed to find the basic pattern of flow of light i.e. the direction in which the image is getting darker or lighter, represented by an arrow.
- c. The image is broken into squares of dimension 16\*16 and then replacing each square with a single arrow that holds the

- majority within the square.
  - d. Replacing all the squares within the image produces a basic facial structure.
- 3 Handling faces looking in different directions.
- a. The core of this algorithm is to detect 68 specific points called landmarks that exist on every face.
  - b. These landmarks comprise of points such as each eye outside edge, each eyebrow inner edge etc.
  - c. Detection of 68 landmarks from a face is take place, a trained machine learning algorithm can detect these 68 specific landmarks on any face.

**B. Text Detection**

Messages in scenes and recordings contain vital and valuable information however it not open by visually impaired. Utilizing content acknowledgment technology, help the visually impaired individual for find explicit shops, money and so forth. At first, a picture of the regular habitat is caught when the client clicks a film board for content preparing. The normal scene like promotion board, road board and so forth data of content from pictures is perceived by utilizing the Optical Character Recognition(OCR). The main innovation Optical Character Recognition (OCR) empowers content extraction from printed archive. The OCR is work dependent on Google's Tesseract Application Programming Interface(API). It utilized English and numerical inscription acknowledgment based strategy. In the wake of changing over to paired picture then just caption data is recognized in the content lines. Because of the distinctive enlightenment condition and lighting, abnormal state commotions are available at caption region. It result lower acknowledgment rate. Versatile binarization strategy is an answer for the above indicated issue. In settled binarization, one limit esteem is settled for the whole picture. In versatile thresholding, edge is pick as for the variety in pictures. The calculation comprises of following advances:

- Initially image is loaded for processing and the textual area is identified.
- It detect MSER regions and uses canny edge detector to further segment text.
- Some of connected components are removed by region properties.
- Tesseract OCR then filters character candidates using connected component analysis.
- Different languages characters have similar stroke width and thickness throughout. Then it filters character candidates using stroke width image.
- Bounding boxes encloses text regions are determined and OCR is applied on text regions.
- Image is now transformed to text document and finally converted to audio file using text to speech conversion software and sound at earphone.

The preparation of Tesseract process incorporates the accompanying advances:

- Box record creation: Training document contains every one of the letters. Jumping square shape's directions of the each letter is additionally connected with it.
- Possible characters age is done: learning dataset is built and highlight age is conceivable.
- Cluster prototyping is made: Outline include genius to type of all character is created.
- Directed non-cyclic word diagrams are created: frequently showing up words lexicon is enlisted.

**C. Docker Image & Quantization**

A model is quantized it implies it very well may be fit in to any little gadget by making littler. Quantized Model methods make it littler to fit on a little gadget. Raspberry PI like little gadget has constrained intensity of calculation and extremely little memory is existed. Neural system preparing is happen by giving numerous modest pushes to the loads, and utilizing these drifting point accuracy to work in little augmentations. Running interface of pre-prepared model is distinctive in nature. The clamor present in system can without much of a stretch adapt by utilizing profound neural system. For preparing neural system utilized and it heavily devour circle space. For instance, 300 MB gliding point design is utilized by AlexNet. The neural associations in neural system has take distinctive size. In 32 bit drifting point, hubs and loads in edges of neural system is put away. By busing idea of quantization, can lessen the document measure by contracting and putting away it in min and max for each layer, at that point pressure happen by believer coast an incentive to 8 bit whole number. The span of record is diminished up to 80%.

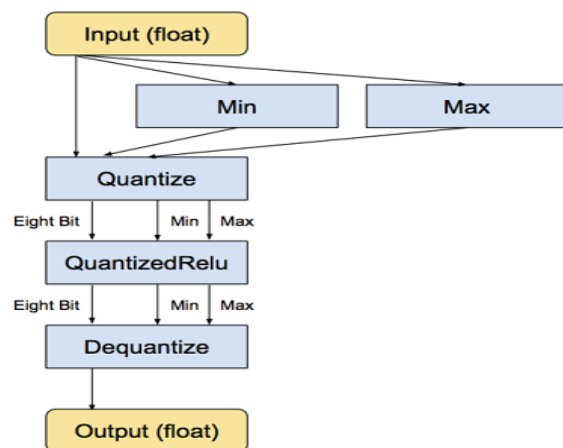


Figure 4.4: Quantization Model



The preparation procedure is extremely hard procedure in any model, it take too long and made critical dimension of intricacy. For settling this issue, docker picture idea is utilized. Utilizing the docker picture the framework trouble for preparing is incredibly diminished. Docker picture produce a basic method for preparing. procedure of preparing a model is pointlessly hard to disentangle the procedure, made a docker picture would make it simple to prepare. In holder, framework previews are available and it is effectively reproduce when the prerequisite is arised in condition. Reproducible nature is a critical component in docker. The pictures are put away in extremely little size and is broadly accessible in compartment so when a necessity of retraining happen, it is effectively done by docker idea.

## V. RESULT

By the effective programming in the module it perceives the feed faces and objects as it required. Sound yield through the headphone. The primary motivation behind this model is to help daze people by controlling them utilizing this framework plan. It perceives the face and text, then people will be distinguished utilizing face and content acknowledgment highlights. It gives the examined and perceived pictures as sound yield to help what's more, direct the visually impaired individual. It is uniquely intended to daze route reason.



Figure 5.1: Hardware Output



Figure 5.2: Face Recognition Output



Figure 5.2: Text Detection Output

## VI. CONCLUSION

Smart vision is effectively actualized and tried. It give gift of vision to person who has restrictions in vision. Keen vision assists the visually impaired individuals for perceive the known individuals in the encompassing condition with in a little deferral. This framework have the ability of lessening the mistake rate in the acknowledgment procedure. Framework has proficiently group the live individuals by utilizing the technique for eye angle proportion. Utilizing this methodology a live face is effectively separated from wax statue and visuals present in the promotion sheets. The content perusing from the common scene is additionally effectively executed by utilizing tesseract API. Messages in scenes and recordings contain critical and valuable data yet it not available by visually impaired. Utilizing content acknowledgment innovation, help the visually impaired individual for find explicit shops, cash and so on. By utilizing the idea of quantization and docker, speed in preparing and retraining segment is enormously decreased. In future, transport ID framework utilizing RFID tag and object identification utilizing yolo calculation are proposed.

## REFERENCES

1. Mandeep Kaur and Jasjit Kaur (2017) "Review of Face Recognition Techniques." In: International Journal of Computer Applications (0975 – 8887) Volume 164 – No 6
2. Zhang W. and Guo Y. (2000) "Feature-Based Face Recognition: Neural Network Using Recognition-by-Recall" In: Mizoguchi R., Slaney J. (eds) PRICAI 2000 Topics in Artificial Intelligence. PRICAI 2000. Lecture Notes in Computer Science, vol 1886. Springer, Berlin, Heidelberg
3. Chucal, Yi, Y. Tian, Areis. Arditi, "Portable Camera-Based-Assistive-Text-and-Product Label Reading From Hand-Held Objects for Blind Persons", IEEE/ASME TRANSACTIONS ON MECHATRONICS.
4. X. Chen, A. L. Yuille, "Detecting and reading text in natural scenes", Proc. Comput. Vision Pattern Recognition., vol. 2, pp. II-366II-373, 2004

5. Arai, K. and Mardiyanto R (2011) "Comparative Study on Blink Detection and Gaze Estimation Methods for HCI, in Particular, Gabor Filter Utilized Blink Detection Method" In: Proceedings of the 2011 Eighth International Conference on Information Technology: New Generations. ITNG '11, Washington, DC, USA, IEEE Computer Society, 2011, pp. 441–446.
6. Akshata. S. (2016) "Eye Blink Detection Using Adaboost Approach and Morphological Operation" In: International journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol 5, Issue 4, April 2016.
7. N. G. Bourbakis and D. Kavradi, "An intelligent assistant for navigation of visually impaired people," in Proceedings of the 2001 IEEE 2nd International Symposium on Bioinformatics and Bioengineering Conference, pp. 230–235, IEEE, 2001.
8. G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind," Applied Soft Computing Journal, vol. 7, no. 1, pp. 257–264, 2007.