Journal of Science, Computing and Engineering Research (JSCER)
Volume-X, Issue- X, Jan-Feb 2022.
DOI: https://doi.org/10.46379/jscer.2022.030101

Smart glasses robot for blind people using raspberry-pi and python

[1]Aisha Abdulatif [2]Sara Mohammed [3]Alaa Mohammed [4]Hind Mohammed [5]Dr.A.Rajalingam

[1,2,3,4,5]Electrical Engineering, University of Technology and Applied Sciences-Shinas, Sultanate of Oman

Article Information

Received	:	January 17 2022
Revised	:	January 31 2022
Accepted	:	February 14 2022
Published	:	February 22 2022

Corresponding Author:
Aisha Abdulatif
Email: Awx.812@gmail.com

Abstract— Visual impairment or vision loss is one of the most important problems that some people suffer from, which is their inability to see what them is around. The blind face several problems, the most important of which is not knowing what is going on around them. Several means have been made to help them, including the use of a stick, but it is not ideal for performing some tasks that the blind person needs. With the development of technology, we can have helped them. This research paper aims to help blind people of all classes accomplish their daily tasks. Smart glasses help by using a camera to identify known and unknown people and detect objects. The blind also faces the problem of reading, so these glasses scan any text and convert it into sound for the blind to hear through a headset. The effort of this model is based on several algorithms, including : (YOLO) to identify objects, (Harr-Cascade) and (LBPH) to identify faces and (OCR) to translate the text. Also, the paper is implemented by using Python programming and Raspberry PI.

Keywords: *Haar-cascade, LBPH, OCR, Python, Raspberry PI, Robot, YOLO.*

Citation: Aisha Abdulatif, Sara Mohammed, Alaa Mohammed, Hind Mohammed and Dr.A.Rajalingam, "Smart glasses robot for blind people using raspberry-pi and python", Journal of Science, Computing and Engineering Research, 3(1), 237-244, 2022.

I. INTRODUCTION

A physical disability has affected the lives of many people all over the world. One of these disabilities that has affected people the most is blindness. According to the World Health Organization, they estimated the number of blind people in the world for the year 2020 to be 43 million [1]. As for the latest statistic in the Sultanate, which was in 2010, according to the Al-Noor Association for the Blind, it estimated the number of blind people in the Sultanate to 21,000 [2]. Thus, the visually impaired lives in a narrow and limited world because of disability and would like if he could get rid of him and go out to the world of sighted people.

But technology has been able to help blind people in their daily lives. The robot smart glasses for the blind were an ideal solution in our time instead of using the traditional stick, by using artificial intelligence and image processing. Our robot is implemented with the following features, recognizing different faces using Haar-cascade algorithm, identifying the person using LBPH algorithm, and identifying the different objects using YOLO algorithm.

Moreover, our robot can read text in the book using OCR algorithm and detected text will be transformed to audio so that blind people can hear the audio. Our robot is implemented using python programming and Raspberry-Pi. Without a doubt, this research will be a turning point in the world of blind people using technology.

In our paper we have studied various artificial intelligence, machine learning and deep learning algorithms, and they are described as follows:

II. FACE DETECTION USING HAAR-CASCADE CLASSIFIER

Much research had been done on Haar cascading techniques [3][4][5]. It is an Object Detection Algorithm used to identify faces in an image or a real time video. The algorithm uses edge or line detection features. This algorithm has four steps:

A. Haar Feature Selection:

Objects are classified on very simple features (Haar filters). In Figure 1, an example of these features is a 2-rectangle feature, defined as the difference of the sum of pixels of area inside the rectangle, which can be any position and scale within the original image [4][5].

B. Integral Image Representation:

The Value of any point in an Integral Image, is the amount of the relative multitude of pixels above and left of that point [26]. In figure 2, an Integral Image can be determined effectively in one pass over the picture [6].

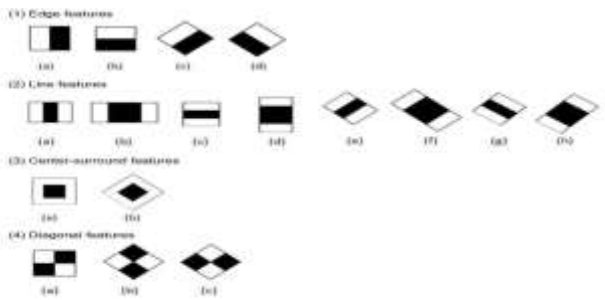

Figure :1 Feature Selection

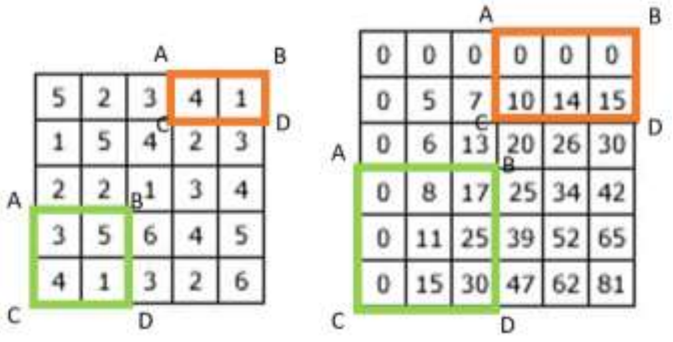

Figure: 2 Integral Image

C. Adaboost Training:

For a window of 24x24 pixels, there can be around 162,336 potential highlights that would be pricey to assess. Thus, in figure 3 , AdaBoost calculation is utilized to prepare the classifier with hands down the best elements [7][9][20].

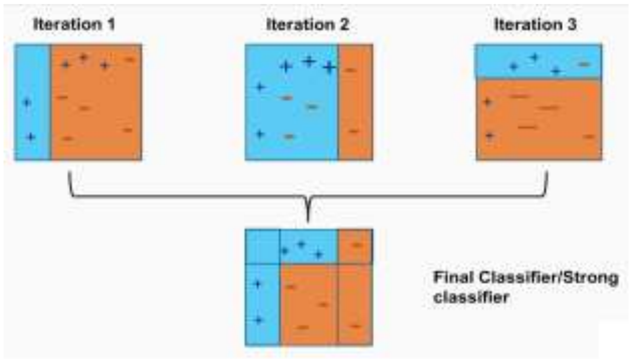

Figure: 3 AdaBoost classifier

D. Cascade Classifier Architecture:

A cascade classifier refers to the concatenation of several classifiers arranged in successive order. It makes large numbers of small decisions as to whether it's the object or not. In figure 4, the structure of the cascade classifier is of a degenerate decision tree [8].

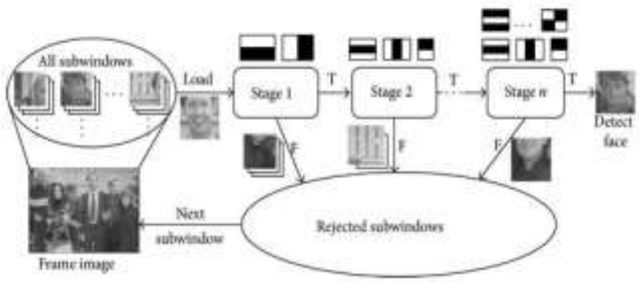

Figure: 4 Cascade Classifier

III. FEATURE EXTRACTION FOR RECOGNITION USING LBPH

The Local Binary Pattern Histogram (LBPH) algorithm is a face recognition algorithm based on a local binary operator, designed to recognize both the side and front face of a human [10]. In figure5, the LBPH algorithm makes use of 4 parameters which is Radius, Neighbours, Grid Y and Grid X. Divide it into regions of same height and width resulting in m x m dimension for every region [24].

Figure: 5 The Face image divided into blocks

Local binary operator is used for every region. The LBP operator is defined in window of 3x3.

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c)$$

Equation: 1 LBP Calculation

Using median pixel value as threshold, it compares a pixel to its 8 closest pixels using this function [25].

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Equation: 2 Value of neighbor

If the value of neighbor is greater than or equal to the central value, it is set as 1 otherwise it is set as 0. Thus, we obtain a total of 8 binary values from the 8 neighbors. After combining these values, we get an 8bit binary number which is translated to decimal number for our convenience. This decimal number is called the pixel LBP value and its range is 0-255.

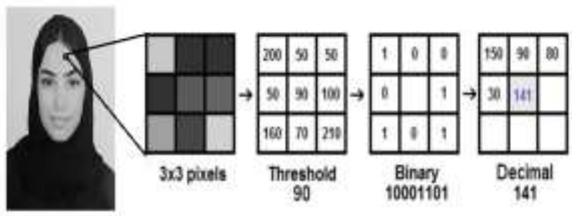

Figure: 6 Local binary patterns

The algorithm was improved to use different number of radius and neighbors, now it was known as circular LBP [11] [14].

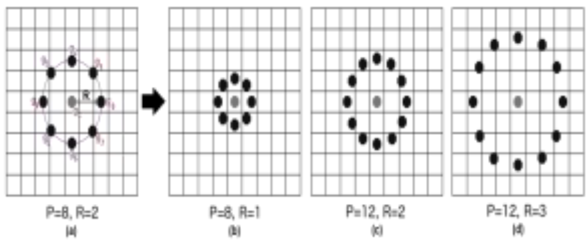

Figure: 7 Circular LBP

After creation of histogram for each region all the histograms are merged to form a single histogram, and this is known as feature vector of the image [10].

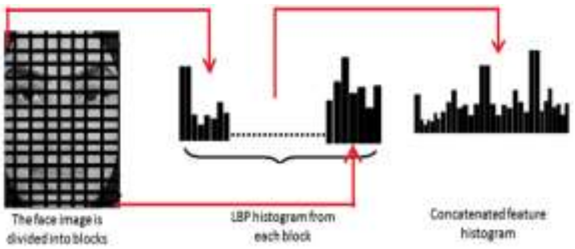

Figure: 8 LBP Histogram equalization

Figure: 9 Test Image Output

As an output we get an ID of the image from the database if the test image is recognized.

IV. OBJECT DETECTION USING YOLO

The expression 'You Only Look Once' is abbreviated as YOLO. This is an algorithm for detecting and recognizing different items in a photograph (in real-time). Object detection in YOLO is done as a regression problem, and the identified photos' class probabilities are provided. Convolutional neural networks (CNN) are used in the YOLO method to recognize objects in real time. To identify objects, the approach just takes a single forward propagation through a neural network, as the name indicates. This indicates that a single algorithm run is used to forecast the whole picture. The CNN is used to forecast multiple bounding boxes and class probabilities at the same time [12] [13]. YOLO algorithm works using the following three techniques:

A. Residual blocks

The picture is first separated into many grids. The dimensions of each grid are S x S. The graphic below demonstrates how a grid is created from an input image. There are several grid cells of identical size in the image above. Objects that occur within grid cells will be detected by each grid cell. If an item center emerges within a certain grid cell, for example, that cell will be responsible for detecting it [13].

Figure:10 Residual blocks

B. Bounding box regression

A bounding box is an outline that draws attention to a certain item in a picture. The following properties are present in every bounding box in the image:

- Length (bw)
- a certain height (bh)
- The letter c stands for class (for example, person, automobile, traffic light, etc.).
- Center of the bounding box (bx,by)

A bounding box is illustrated in the figure below. A yellow outline has been used to depict the bounding box [13].

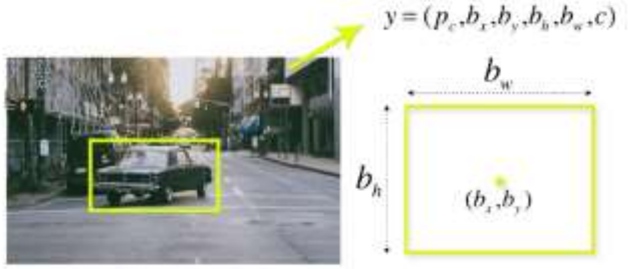

Figure: 11 Bounding box regression

To forecast the height, breadth, center, and class of objects, YOLO use a single bounding box regression. The likelihood of an object occurring in the bounding box is represented in the image above.

C. Intersection over union (IOU)

The concept of intersection over union (IOU) illustrates how boxes overlap in object detection. IOU is used by YOLO to create an output box that properly surrounds the items. The bounding boxes and their confidence ratings are predicted by each grid cell. If the anticipated and real bounding boxes are identical, the IOU is 1. This approach removes bounding boxes that aren't the same size as the actual box [13].

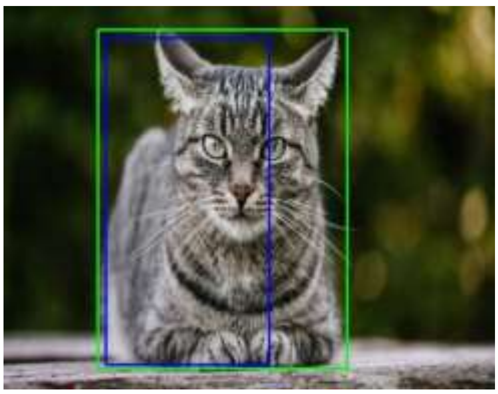

Figure: 12 Intersection over union (IOU)

There are two bounding boxes in the image above, one in green and the other in blue. The blue box represents the anticipated box, and the green box represents the actual box. YOLO makes sure the two boundary boxes are the same size.
The graphic below depicts how the three techniques are combined to generate the final detection findings [13].

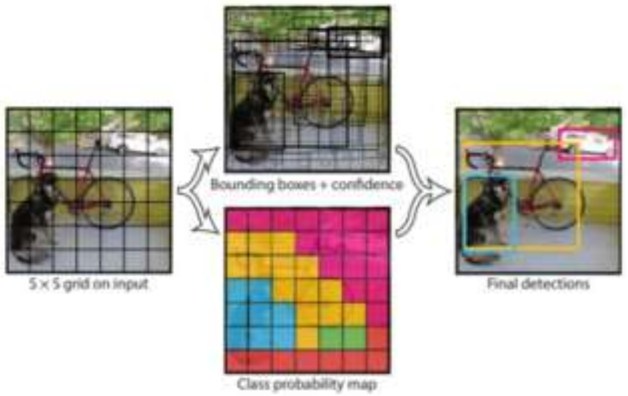

Figure: 13 Final detection results

The picture is first subdivided into grid cells. B bounding boxes are forecasted in each grid cell, along with their confidence scores [13]. To determine the class of each item, the cells estimate the class probability. We can see at least three types of items, for example: a vehicle, a dog, and a bicycle. A single convolutional neural network is used to make all the predictions at the same time. The predicted bounding boxes are equivalent to the true boxes of the objects when intersection over union is used. This phenomenon gets rid of any extra bounding boxes that don't fit the objects' properties (like height and width). The final detection will be made up of distinct bounding boxes that exactly suit the objects. The pink bounding box, for example, surrounds the automobile, whereas the yellow bounding box surrounds the bicycle. The blue bounding box has been used to highlight the dog.

V. TEXT TO SPEECH CONVERSION USING OCR

The electronic translation of typed, handwritten, or printed text images into machine-encoded text is known as optical character recognition (OCR) [15] [16] [17] [21]. With OCR, many paper-based documents in a variety of languages and formats may be turned into machine-readable text, making previously inaccessible material accessible to anybody with a single click. This problem is difficult to tackle since there are so many different fonts and ways to write a single character. Before choosing an OCR method, the picture must first be preprocessed so that it can be "read." OCR software frequently "pre-processes" pictures to improve recognition possibilities. This technique includes the following features [21] [22]:

i) De-skew

If the document was not properly aligned when scanned, it may need to be slanted a few degrees clockwise or counterclockwise to make fully horizontal or vertical text lines.

ii) Despeckle

Remove all positive and negative marks while also smoothing down the borders.

iii) Binarization

Convert a picture to black-and-white (sometimes known as a "binary image" due to the two hues). The binarization job is used to identify text (or any other needed picture element) from the backdrop in a simple and precise manner.

iv) Line removal

Removes non-glyph boxes and lines.

v) Layout analysis or "zoning"

Columns, paragraphs, captions, and other elements are identified as blocks. In multi-column layouts and tables, this is especially beneficial

vi) Line and word detection

Sets a baseline for word and character forms, and separates words as needed.

vii) Script recognition

Because the script in multilingual documents might change at the word level, script identification is required before the appropriate OCR can be used to handle the script.

viii) Character isolation or "segmentation"

Various characters linked by picture artifacts should be split, and single characters fragmented into several artifact-based fragments should be linked for OCR characters.

ix) Normalization

Scale and aspect ratio should be normalized.

D. A. OCR Extraction of Features

In OCR, there are two approaches for extracting features: The first technique uses a feature detection algorithm that evaluates a character's lines and strokes to characterize it. Pattern recognition, in the second method, identifies the full character. A line of text can be identified by looking for white pixel rows with black pixels in between. In the same way, we can tell where a character begins and ends [18] [19] [21] [22]. The depiction of these methods is depicted in the following images:

Figure: 14 Method 1 – Feature detection

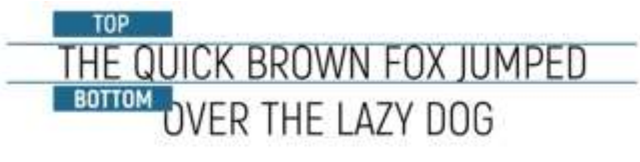

Figure: 15 Method 2 – Pattern recognition on a row of text

Figure: 16 Method 3 – Pattern recognition on a single character

Then, as illustrated in the following graphic, we convert the character's image into a binary matrix, where white pixels are 0s and black pixels are 1s [21].

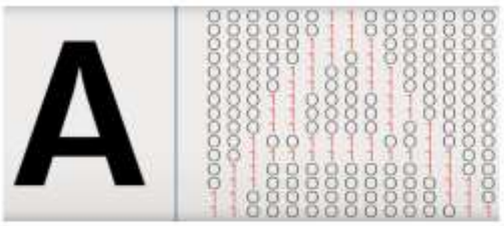

Figure: 17 Sample of binary matrix

The distance from the matrix's center to the farthest 1 may then be calculated using the distance formula [22].

$$d = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}$$

Equation: 3 Distance formula

Then we make a circle with that radius and divide it into smaller portions. At this point, the algorithm compares each subsection to a database of matrices representing characters in various fonts to determine the character with which it shares the most statistical similarities. It makes it simple to convert printed media to digital by doing so for each line and character [15] [16] [21] [22].

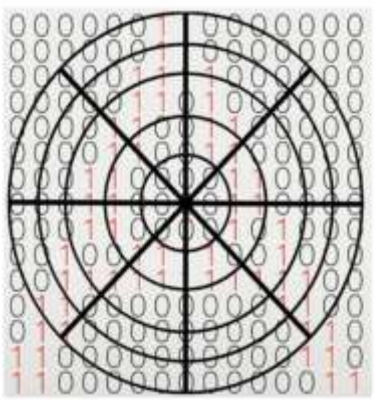

Figure: 18 Compare each subsection against the matrix database

B. Post processing

If the output is constrained by a vocabulary, post-processing OCR accuracy can be enhanced (a list of words permitted in a document). For example, this may be the whole English vocabulary or a more technical lexicon for a certain subject [19].

If the text contains terms that are not in the lexicon, such as proper nouns, this strategy may be less effective. Fortunately, there are free online OCR libraries accessible to boost accuracy [21]. The Tesseract library uses its dictionary to manage character segmentation [23]. The output stream can be a single string or a character file, but more powerful OCR systems save the page structure and, for example, generate a PDF with both the original picture pages and a searchable textual image [21].

C. Correction of errors

By recording those particular words have been encountered together, the "near neighbor analysis" can utilize frequencies for co-occurrence to remedy inaccuracies. In English, for example, "Washington, D.C." is more common than "Washington DOC" [21].

D. Grammar

Grammar may also aid in determining the language being scanned, for example, determining if a word is a verb or a noun, which improves accuracy. The Lowenstein Distance technique is frequently used in OCR post-processing to improve OCR API results [21] [18].

VI. PROPOSED METHOD

Smart glasses robot has three important tasks as to identify known and unknown people, to identify objects, and to read text from book. Figure 19, 20 and 21 shows the proposed model in detail. Our main contribution is to develop the above functionalities using raspberry-pi, camera, headphones, glasses, and python programming. We trained our robot to detect the faces using Haar-cascade classifier algorithm, to identify the known person using feature extraction for recognition using LBPH algorithm, to identify the object using YOLO algorithm, and to read the text using OCR algorithm. All the inputs are given to our robot using camera. All the results are converted to audio and given through headphone to the blind people.

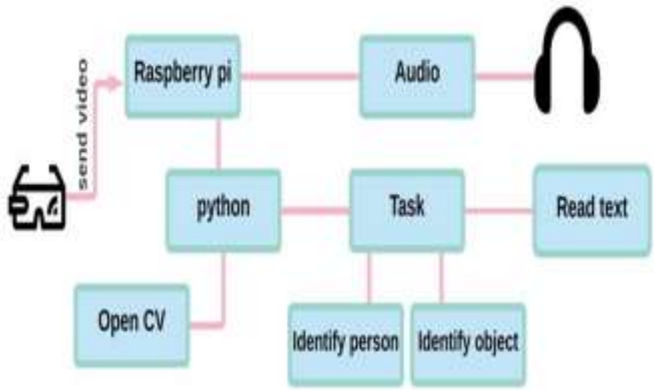

Figure: 19 Block Diagram of Proposed Model

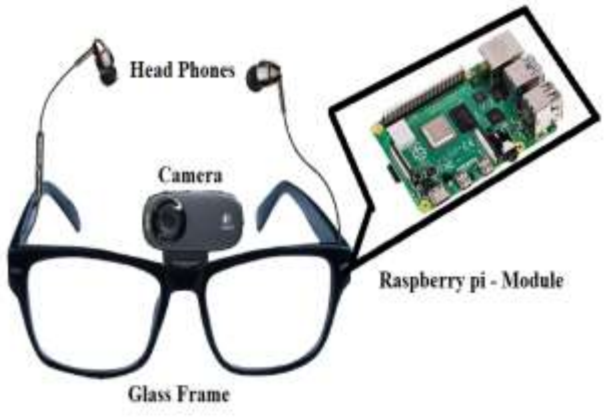

Figure: 20 Proposed Prototype

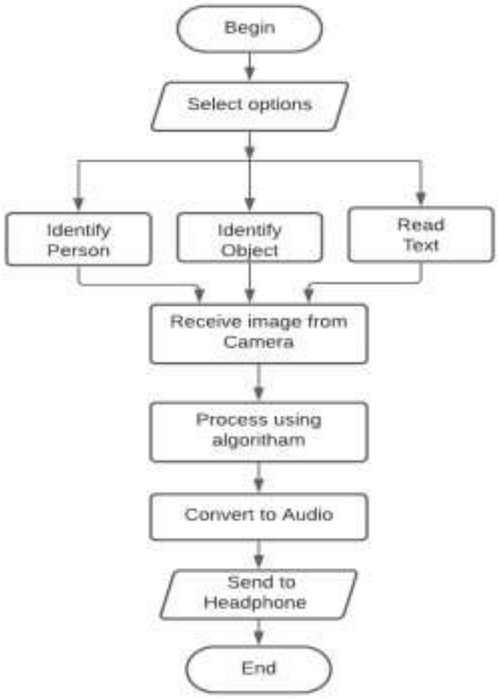

Figure: 21 Proposed Model Flow chart

Figure 22 shows the flow of Haar-cascade and LBPH. Video/Image is given through the camera to this module and Haar-cascade algorithm helps to detect the number of faces in the given input and LBPH algorithm helps to identify the known and unknown person from the existing database. Result is converted into audio/voice format, and it sends to headphone, so that the blind people can identify the persons in and around.

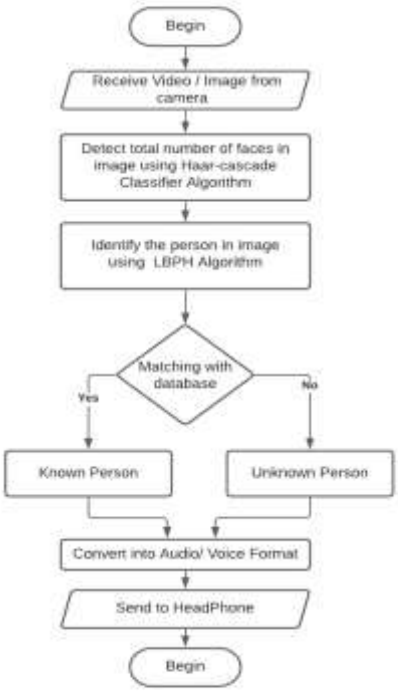

Figure: 22 Haar-Cascade and LBPH Flow chart

Figure 23 shows the flow of YOLO. Video/Image is given through the camera to this module and YOLO algorithm helps to identify the number of objects and its name from the existing database. Result is converted into audio/voice format, and it sends to headphone, so that the blind people can identify the objects in and around.

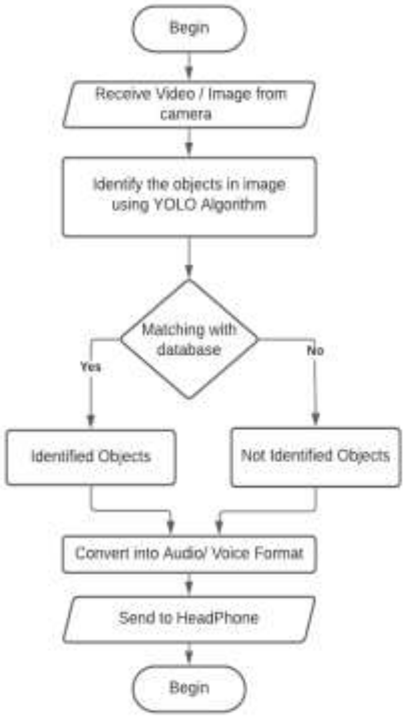

Figure: 23 YOLO Flow chart

Figure 24 shows the flow of OCR. Video/Image is given through the camera to this module and OCR algorithm helps to read the texts and identified words are converted into audio/voice format and it sends to headphone, so that the blind people can read the words from book.

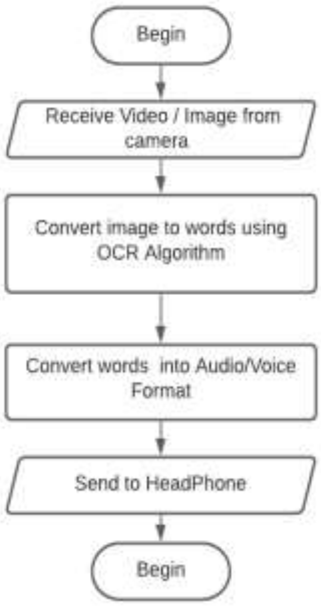

Figure: 24 OCR Flow chart

VII. RESULTS

Table I gives the results of proposed model. Here the dataset is number of faces identified and accuracy rate is in the form of percentage.

Table I: Number of Faces and Person Identified

No. of faces in front of camera	Execution Time (sec)	No. of faces Identified	Accuracy (%)	Known person Identified	Unknown person Identified	No. of faces not Identified
4	1.492	4	100	4	0	0
8	2.193	7	87.5	4	3	1
12	2.949	9	75	5	4	3
16	3.261	11	68.7	5	5	6

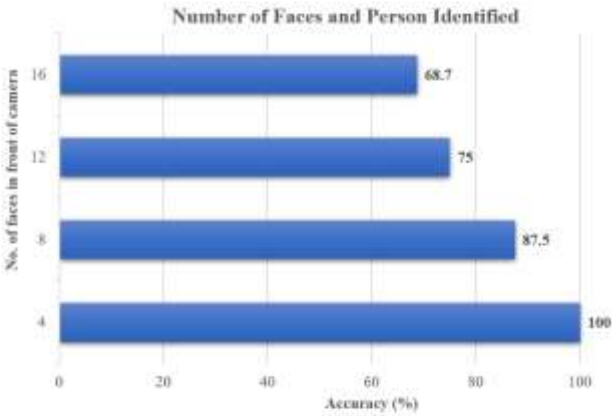

Figure: 25 Number of Faces and Person Identified

Table II gives the results of proposed model. Here the dataset is number of objects identified and accuracy rate is in the form of percentage.

Table II: Number of objects Identified

No. of objects in front of camera	Execution Time (sec)	No. of objects Identified	Accuracy (%)
4	0.163	4	100
8	0.257	8	100
12	0.385	10	83.3
16	0.442	13	81.2

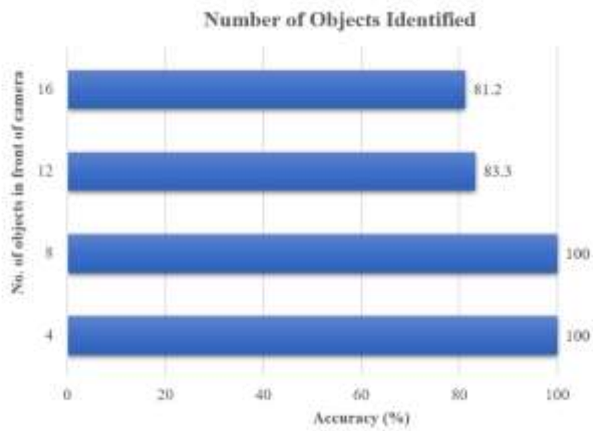

Figure: 26 Number of Objects Identified

Table III gives the results of proposed model. Here the dataset is number of words identified and accuracy rate is in the form of percentage.

Table III: Number of words Identified

No. of words in front of camera	Execution Time (sec)	No. of words Identified	Accuracy (%)
100	2.387	94	94
200	3.984	174	87
300	5.123	235	78.3
400	5.982	318	79.5

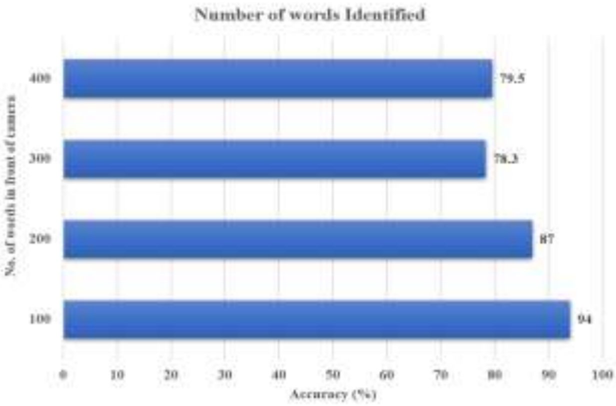

Figure: 27 Number of Objects Identified

Based on the above results, we achieved the higher accuracy when number of faces, objects and words are less. And we achieved the lesser accuracy when number of faces, objects and words are more in the video frame. Execution time of our algorithms is very less. So, processing speed is high. So, in future to improve the accuracy, we are recommending adding some machine and deep learning algorithm to process should be given to out proposed robot.

VIII CONCLUSION

Smart robot glasses are one of an important technology which can help many blind people and offers a lot of significant solution for their problems, this type of robot glasses will help many blind people to identify the people, objects and read text. We analyzed Haar-cascade and LBPH algorithms together used to identify faces, YOLO algorithm used to detect objects and the OCR algorithm used to identify and convert the texts into the audio. These smart glasses robot can detect faces and objects with high accuracy with the help of these algorithms. Finally, after detection and recognition of the face and objects, and texts with a good accuracy for the given dataset, the name of the recognized person and object is converted into speech and given as output to the headphone.

REFERENCES

[1]. John Elflein, " Number of blind people worldwide in 2020, by age and gender", 2021.
[2]. Alroya newspaper, "21 thousand blind people in the sultanate, according to 2010 statistics", ALNOOR Association for the blinds ,2015.

[3]. K.C. Paul , S Aslan , An improved real-time face recognition system at low resolution based on local binary pattern histogram algorithm and CLAHE, Optics Photonics J. 11 (April 2021) 63–78.

[4]. F.M. Javed Mehedi Shamrat, Anup Majumder, Probal Roy Antu, Saykot Kumar Bar- mon, Itisha Nowrin, Rumesh Ranjan "Human Face Recognition Applying Haar Cas- cade Classifier "International Conference on Pervasive Computing and Social Net- working, Salem, Tamil Nadu, India, 19-20, March 2021.

[5]. Arnav Madan , Face recognition using haar cascade classifier, Int. J. Modern Trends Sci. Technol. 07 (01) (January 2021) 85–87.

[6]. Mohamed OUALLA, Dr. Abdelalim SADIQ, Dr. Samir MBARKI, Comparative Study Of The Methods Using Haar-Like Features , International Journal Of Engineering Sciences & Research Technology 4(4) (April 2015) 35–43.

[7]. Shuqiong Wu and Hiroshi Nagahashi, Analysis of Generalization Ability for Different AdaBoost Variants Based on Classification and Regression Trees , Journal of Electrical and Computer Engineering-Volume 2015, Article ID 835357, 17 pages.

[8]. Sarala A. Dabhade & Prof. Mrunal S. Bewoor , Real Time Face Detection and Recognition using Haar - Based Cascade Classifier and Principal Component Analysis" International Journal of Computer Science and Management Research Vol 1 Issue 1 Aug 2012.

[9]. Wen-Chang Cheng and Ding-Mao Jhan, A Cascade Classifier Using Adaboost Algorithm and Support Vector Machine for Pedestrian Detection" IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. 32, No. 7, pp. 1239 - 1258, 2010.

[10]. M. A. Abuzneid and A. Mahmood, Enhanced Human Face Recognition Using LBPHDescriptor, Multi-KNN, and Back-Propagation Neural Network 2018 IEEE Access,VOLUME 6, 2018 Page No:20641-20651.

[11]. X. Zhao and S. Zhang, ''Facial expression recognition based on local binary patterns and kernel discriminant Isomap,'' MDPI Sensors, vol. 11, no. 10, pp. 9573–9588, 2011.

[12]. Jing Tao, Hongbo Wang, Xinyu Zhang, Xiaoyu Li and Huawei Yang, An Object Detection System Based on YOLO in Traffic Scene, International Conference on Computer Science and Network Technology-IEEE 2017 Pages:315-319.

[13]. Onesmus Mbaabu "Introduction to YOLO Algorithm for Object Detection", April 15, 2021.

[14]. Md. Abdur Rahim, Md. Najmul Hossain, T. Wahid, Md. Shafiul Azam, "Face recognition using Local Binary Patterns (LBP)", Global Journal of Computer Science and Technology Graphics &Vision,, Global Journals Inc. (USA), Volume 13 Issue 4 Version 1.0 Year 2013.

[15]. Wang, T., Wu, D. J., Coates, A., & Ng, A. Y. (2012, November). End-to-end text recognition with convolutional neural networks. In Pattern Recognition (ICPR), 2012 21st International Conference on (pp. 3304-3308). IEEE.

[16]. Koo, H. I., & Kim, D. H. (2013). Scene text detection via connected component clustering and nontext filtering. IEEE transactions on image processing, 22(6), 2296-2305.

[17]. Bissacco, A., Cummins, M., Netzer, Y., & Neven, H. (2013). Photoocr: Reading text in uncontrolled conditions. In Proceedings of the IEEE International Conference on Computer Vision (pp. 785-792).

[18]. Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2014). Robust text detection in natural scene images. IEEE transactions on pattern analysis and machine intelligence, 36(5), 970- 983

[19]. Neumann, L., & Matas, J. (2012, June). Real-time scene text localization and recognition. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 3538- 3545). IEEE.

[20]. P. Viola and M. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a DetectorCascade", In Proc. of NIPS01, 2001.

[21]. Forough Karandish," The Comprehensive Guide to Optical Character Recognition (OCR)"

[22]. Nikita Kotwal,Ashlesh Sheth,Gauri Unnithan and Gauri Unnithan, Optical Character Recognition using Tesseract Engine - International Journal of Engineering Research & Technology (IJERT), Vol. 10 Issue 09, September-2021 Pages:381-385.

[23]. S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S.Wong and R.Young. "ICDAR 2003 Robust Reading,Competitions", In 7th International Conference on Document Analysis and Recognition2003.

[24]. J. CHAO W L, DING J J, LIU J Z. Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection. Signal Processing, 2015,117:1-10.

[25]. Marko Arsenovic, Srdjan Sladojevic, Andras Anderla, Darko Stefanovic, FaceTime Deep Learning Based Face Recognition Attendance System, IEEE 15th International Symposium on Intelligent Systems and Informatics September 14-16, 2017.

[26]. Jun Nishimura and Tadahiro Kuroda," Versatile Recognition Using Haar-Like Feature and Cascaded Classifier", Ieee Sensors Journal, Vol. 10, No. 5, May 2010.