FACE RECOGNITION WITH VOICE APPLICATION

Abstract

Facial recognition may be a biometric identification method that is non-intra-operative and suited for many different kinds of applications. It is necessary for implementing the method with a considerable while and appropriate precision while taking hardware timing into account. The application of a machine learning method for instant facial picture recognition is the main focus of this study. The face recognition software employs algorithms to virtually confirm a person's identification by comparing a digital image taken with a camera to a face print that has been stored. One of the key facial recognition techniques created was the Haar-Cascade approach. The HOG (Histogram of Oriented Gradients) approach has been used to recognize faces and has proven to be quite successful for visual perception. The Eigen feature-based face recognition algorithm is compared to both techniques. Using a Raspberry Pi and its camera, the facial recognition model is put into practice to detect and identify faces in real-time for both the pre-existing databases as well as user-defined databases. After identifying the face, we'll use the libraries (GTTS, Play sound) to connect the machine learning algorithm with the audio. It will be able to detect, identify, and respond to the individual in this way.

Keywords: Haar-Cascade, Trainer, HOG, SVM, Voice Recognition, GTTS.

Authors

D. Madhusudan
Assistant Professor
Department of Electronics and Communication Engineering
St. Peter’s Engineering College
Maisammaguda, Hyderabad, Telangana, India.
madhusudan.d@stpetershyd.com

Prudhvi Raj Budumuru
Assistant Professor
Department of Electronics and Communication Engineering
Vishnu Institute of Technology
Bhimavaram, India.
prajece2005@gmail.com

I. INTRODUCTION

The process of determining whether a previously detected object is a known or unknown face is known as face recognition. Face detection and recognition are sometimes misconstrued as separate issues. Face Recognition, on the other hand, uses a database of faces to authenticate the input face by determining whether the "face" belongs to a known or unknown person. Finding faces in photographs is a potential computer vision challenge known as face detection. It is an easy problem for humans to solve, and traditional feature-based algorithms like the cascade classifier have done a good job at doing so. The Multi-task Cascade Convolutional Neural Network, or MTCNN for short, is one example. As face recognition became more and more popular among researchers, numerous different algorithms were created; three of them are well-researched in the face recognition literature.

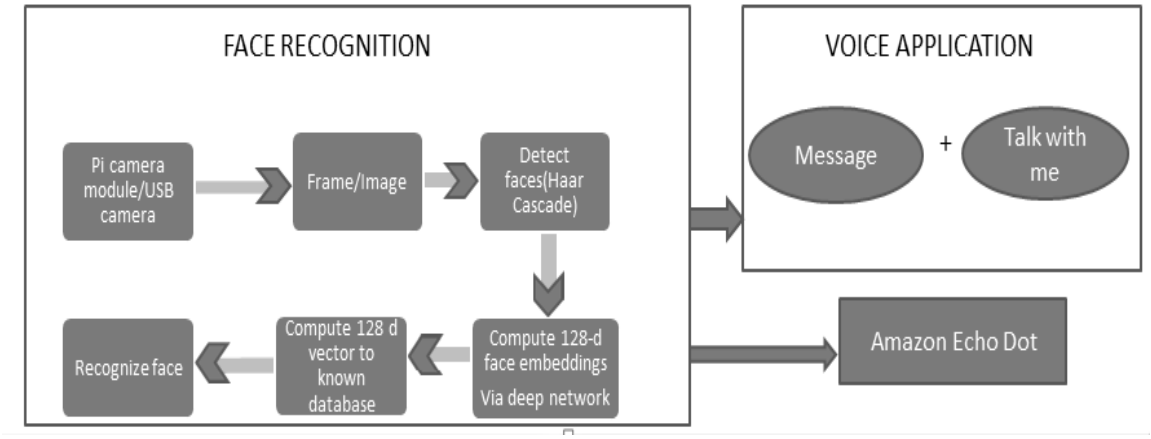

Figure 1: Integration of Face Recognition with Voice Application

The various functional blocks of the project in this system are listed below:

- **USB camera –**It is an electronic component used to capture the image from the user.
- **Frame/Image –** It is a stored file that is taken as a pre-defined input from the usb camera to detect faces of the user.
- **Detect Faces –** Identify and checks whether the image given, actually has a face of a person or not.
- **Compute 128-d face embeddings -** To obtain the embeddings, which are 128-d feature vectors used to characterize each face in an image [1]
- **Compute 128d vector –** To calculate the 128-d feature vectors for all the data samples in the data base.
- **Recognize face –**Verifying data samples from the data set and identify whether the image is present in the dataset or not
- **Message –** Voice message which is given as output after recognizing a face
- **Talk with me –** An application developed to interact with a person
- **Echo Dot –** An electronic speaker embedded with artificial intelligence

II. LITERATURE WORK

The author expressed two main approaches for the face- recognition process the two being geometrical and photometrical. The geometrical is based on the features of the image

that is taken as input whereas the photometrical is based on the view-based analysis. Even though several techniques have been developed, only Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA), and Elastic Bunch Graph Matching (EBGM) have received extensive investigation and study.[2] According to the author, the face identification process uses four fundamental techniques: feature-based, appearance-based, knowledge-based, and template matching. It depends on both structural locations and the images that include the facial as well as non-facial parts that are included in the first technique, which is feature-based. In the second method i.e the appearance based we find that the image is detectable only by the process of statistical analysis and machine learning algorithms. In the third method i.e. the knowledge based we find that the rules are according to the human knowledge such as whether the image has two eyes, one nose etc. In the fourth i.e the template matching we find that the method used a pre-defined set of rules or templates and then compare the given image with the predefined templates. [3] In a geometric-based technique, the classification of faces is done by first extracting and computing the local features of the face, such as the lips, eyes, brows, and nose. Active Appearance Graph Models (AAM) are the most effective geometric-based techniques. [4]. The photometric based technique has different recognition algorithms that are applied to the various methods available to users by using the photo-face database. Two different techniques inside are retrieved by the albedo image and the 3-D facial geometry indicated by the depth of the normal field [5]. The author of this article describes how to perform facial recognition using OpenCV. Face detection is the first step in creating a face recognition program, followed by the extraction of face embeddings from each and every face utilized in deep learning. The model then uses OpenCV to recognize the faces in the pictures after training a face recognition replica on the embedded data [6].

III. STUDY ON VARIOUS APPLICATIONS AND METHODOLOGY

Facial images are used by facial recognition systems to identify persons. Instead, then just verifying the validity of an ID, face recognition systems prove the presence of an official person. The facial recognition technology does not employ ID numbers to distinguish one person from another; instead, it compares the faces of the people directly. Manual assessment is necessary to confirm that the top two matches are in fact different people in order to eliminate duplicates when they are extremely similar to the query face image [7].

Access Control: The size of the population that needs to be identified is typically quite tiny in many access control applications, such as office access or computer logon. Additionally, natural lighting and frontal faces are used to photograph people's faces. Without much user assistance, this application's facial recognition algorithm can operate with excellent accuracy. The examples are as follows.

Face recognition technology is used to track who is in front of a computer terminal continually. It enables the user to log out and close files without closing the terminal window. A pattern hides the work and turns off the keyboard and mouse when the user walks away for a predetermined amount of time. The previous session displays because it was left in its original state when the user returns and is recognized. The attempt by the other user to log on without permission is rejected. [8]

Security: These days, security is of utmost importance to airport staff, airline departments, and passengers. Face recognition technology-based airport security systems have already been put in place at numerous airports all over the world. Every time someone who resembles a known terrorist or suspect enters the airport's security checkpoint, the technology is designed to inform the public safety personnel there. Anyone identified by the system would be subject to additional police investigation. The use of facial recognition technologies in computer security has expanded recently. Users are repeatedly authenticated in order to ensure that the person in front of the computer screen or at a user is the same authorized person who logged in. This prevents someone from changing files or conducting illegal transactions with an unidentified individual when an authorized individual leaves the computer for a brief period of time. research into image databases searching image databases of police bookings, missing children, immigrants, and licensed drivers.[9] Identification verification, banking, e-commerce, employee IDs, passports, national IDs, identifying births, and voter registration are all examples of general identification. [10] *Surveillance:* The implementation of face recognition systems for massive scale surveillance is a difficult undertaking because to factors such public spaces, security monitoring by face recognition system, and other dividers. Some instances of face-based surveillance include the ones that follow. [11].

There are three methods which can be used for the face recogniton.1) Holistic matching methods, 2) Feature- based 3) Hybrid methods. [05]. In the holistic approach entire face is taken as an input. Eigen faces are a best example for the holistic matching methods [06]. Local features such as eyes and mouthand nose are extracted and their addresses i.e the locations are aligned into the structural classifier [06].

Hybrid is a combination of both holistic and also the feature extraction techniques. Any 3d image that is caught is actually an important ingredient used in the hybrid methods [06]. The classical face recognition algorithms are based on the external structure. The eigenface images are projected into a linear subspace [07]. These classical methods are a failure because they fail to represent when a large variation is found and an illumination expression occurs [07]. Artificial neural networks are also used in the face recognition process. They are different when compared to a classical method because they are used to solve a non-linear problem. In order to solve the pattern and identify a human face, a chaotic non-convergent network is used [08]. A non-negative sparse coding to learn facial features using normalized cross-correlation is applied. [08]. The Histogram of oriented Gradients is descriptors that are used in the optimization process and also in the computer vision. They are used in the pattern recognition are recognition of visual objects i.e faces. [09] HOG is reminiscent of edge orientation histogram, SIFT descriptor and shape context. They are calculated in dense cells that are normalized and overlapped with the local contrast histogram in the direction of the image gradient to improve the performance of the detector. Therefore, due to the distribution of local intensity gradients, this feature set is very effective for other types of shape- based objects (ie, face detection), even without any knowledge of the corresponding gradients. To extract the HOG descriptor, first calculate the appearance of the edge direction in the local neighborhood of the image. This means that the image is divided into small connected areas, called cells, and edge-direction histograms are calculated for each. According to whether the gradient is signed or unsigned, the channels of the histogram are distributed from (0 °-180°) or (0°-360°). In order to compensate for the light, the histogram count is normalized by accumulating a measure. The local histogram energy

measurement of the connected area and then using the obtained result to normalize all cells in the block. Finally, the combination of these histograms represents the descriptor HOG. [09]

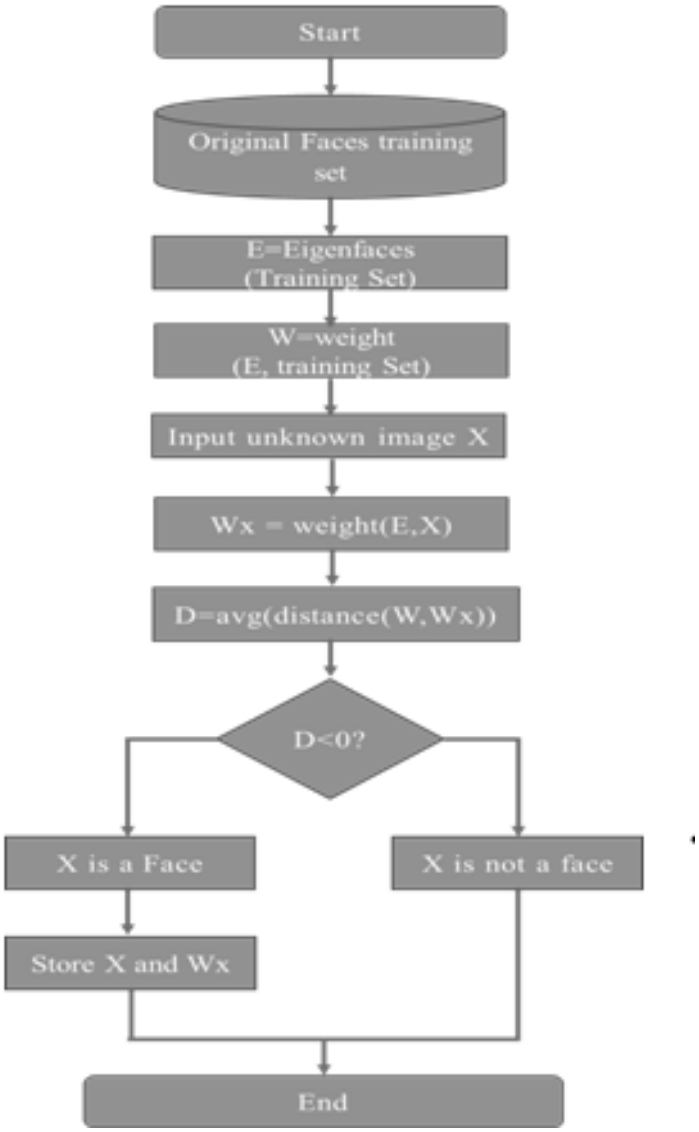

Figure 2: Flow chart of eigenface-based algorithm [2]

In the above flowchart clearly gives the algorithm in detail part with eigen face. The use of computer technology for tasks involving natural (human) languages or speech is made possible by speech recognition technology. The interaction between computers and human (natural) languages is now a major focus of computer science. Data entry for voice recognition systems involves audio input. Speaking through a microphone yields the same outcome as manually typing words on a keyboard. A library of identifiable words, phrases, and sentences is used internally when developing voice recognition software. The software compares spoken audio frequencies with related database information. With AT&T's Bell Labs, the investigation got underway in 1936. Universities and the U.S. government, particularly the military and DARPA, sponsored and carried out the majority of research. [10]

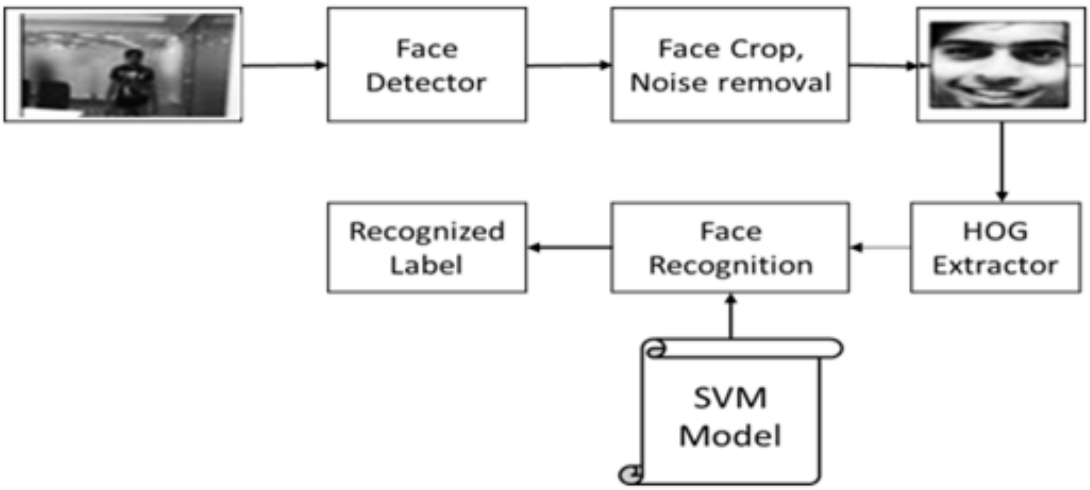

Figure 3: The Histogram of oriented Gradients [12]

How does a computer translate spoken words or voice into data that it can comprehend and act upon? For speech recognition technology, the input is the speaker's voice into a microphone. An analog-to-digital converter converts the electrical signal from the microphone into digital form, which is then stored in memory.[11] The computer compares the speech input to a sample voice format, or voice sample, that has a known meaning to the computer in order to determine the meaning of the voice input. A stream of audio samples are fed into the software as input during speech recognition. Prior to that, a comparable feature must be computed from the incoming audio signal and the speaker database must be available. The database's data is then compared to the feature vectors. This method is comparable to the conventional keyboard-based command inputs. The program contains input templates and makes an effort to match the input given to the computer with these templates. In reality, the finest voice recognizers in the early 1990s reported a 15% error rate on a relatively easy 20,000-word dictation test. Though this can vary widely from speaker to speaker, the error rate has now decreased to as little as 1-2%.[11]

The use of voice recognition is significantly assisting the disabled and handicapped. The performance of speech-recognition technology and contemporary technology in respect to the needs of the disabled population has significantly improved over the past ten years. Among those who benefited most from speech recognition computer technology were blind persons. We employ straightforward devices that read text on early text-based computer screens and translate it into synthetic speech. Additionally, it receives input from verbal commands spoken by a blind person, allowing them to use the computer more effectively and conveniently. The speech recognition software is helpful for persons who have trouble using their hands in addition to blind people.[12][13]

IV. WORKING METHODOLOGY

The below flow chart gives the detailed working methodology of our proposed work from collecting the data set to identification of features for different images.

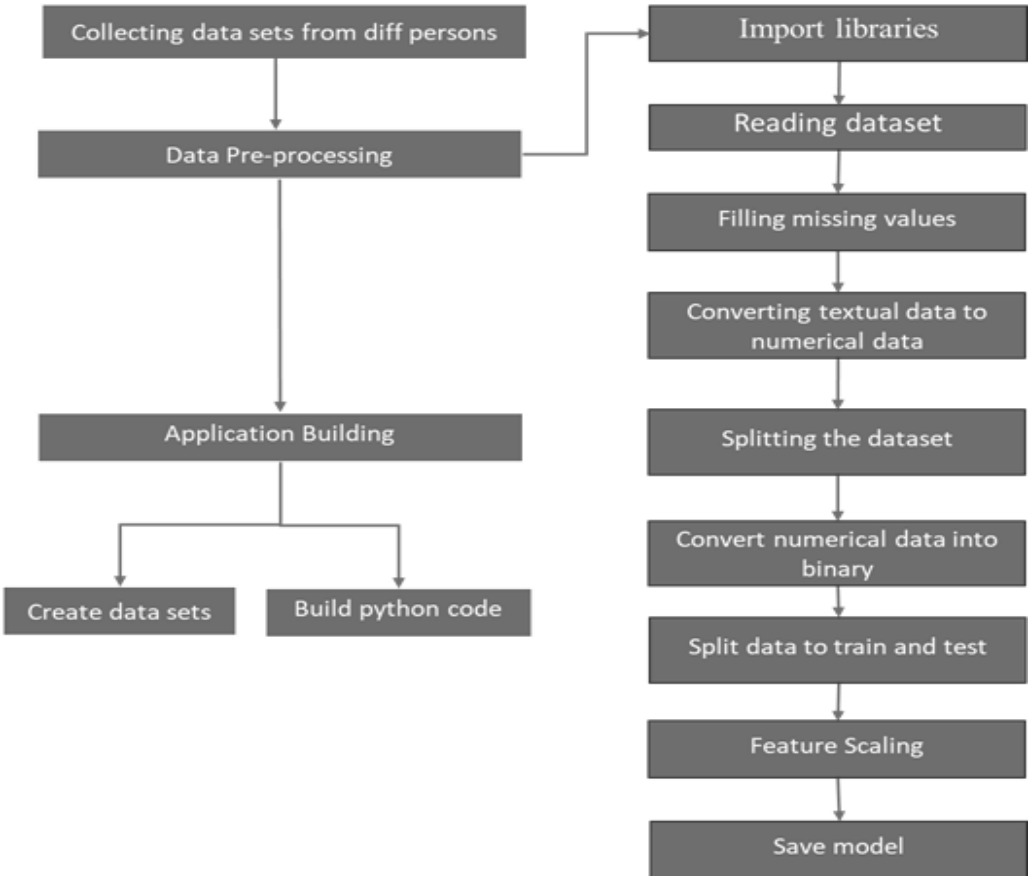

Figure 4: Working methodology

V. IMPLEMENTATION

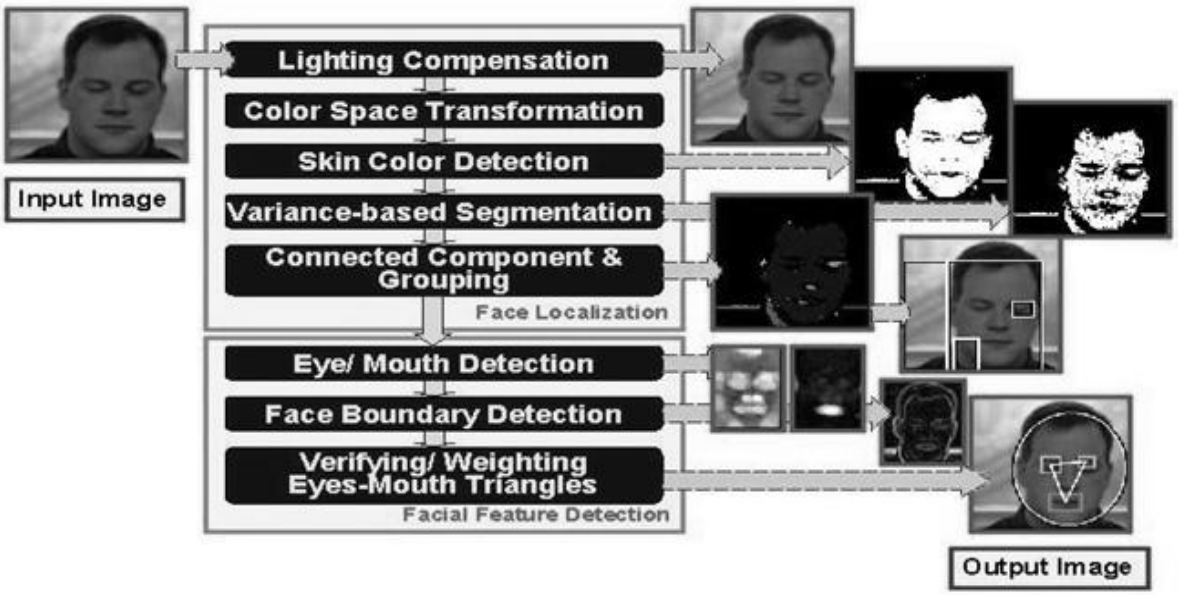

Figure 2: Flow chart for Face detection

VI. WORK FLOW

1. Data Set Creation

- The first step we complete is the creation of a data set.
- The identification of the faces depends on this phase. In this stage, we remove the faces from the image and extract the traits that will subsequently be utilized to identify the faces.
- The detection of multiscale function might be a broad object detection function. It recognizes that since we are calling it on the face cascade. Grayscale is used in the first picture. The multiplier is the second.
- Some faces will appear larger than the faces in the back since they are also closer to the camera. This is offset by the dimensions factor.
- The object detection algorithm employs a moving window. Before announcing that a face has been discovered, Min Neighbors specifies the number of items that have been spotted nearby.

Data Set

Figure 6: Images in DATASET folder

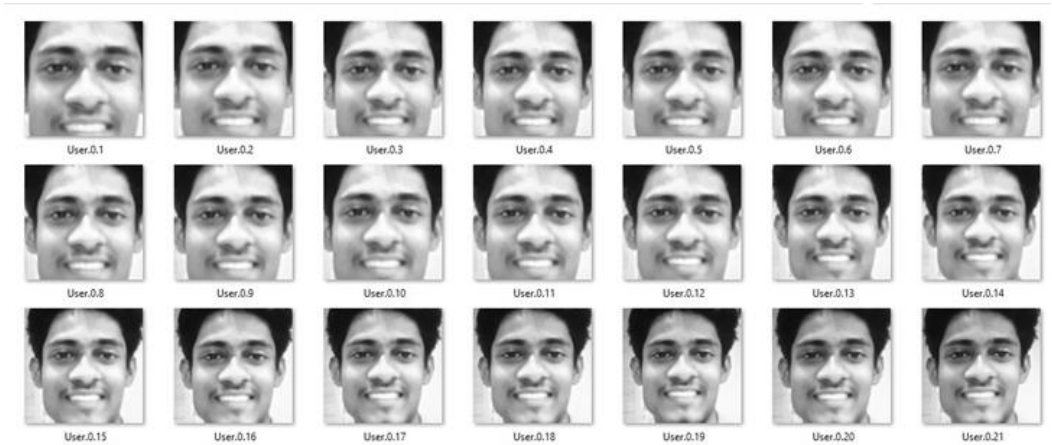

Figure 3: Images in DATASET folder

2. Trainer Process

- TRAINING THE MACHINE is the second step in the facial recognition process.
- Both the HAAR cascade and the training file are used for this. A texturing operator called Local Binary Pattern (LBP) can label each pixel in a picture by thresholding its immediate surroundings and treating the result as a binary number. Let's say we have a grayscale image of a face.
- Think of a small portion of this image as a 3x3 pixel window that contains the intensity of each pixel (0–255). Then, we would like to mandate that the threshold be set using the matrix's central value. The new values from the eight neighbors are defined by this value. We set a fresh binary value for each neighbor of the threshold, the central value.
- We set 0 for values below the brink and 1 for values equal to or higher than the brink. The central value will no longer be present in the matrix; only binary values will. In order to create a replacement binary value (such as 10001101), we want to concatenate each binary value from each location in the matrix line by line.
- Take note that while some authors concatenate the binary values in different ways (such as in a clockwise manner), the end result is always the same.
- The central value of the matrix, which is actually a pixel from the starting image, is then set to this binary value once it has been converted to a decimal value. At the conclusion of this process, we have a new image that more accurately captures the qualities of the initial image.

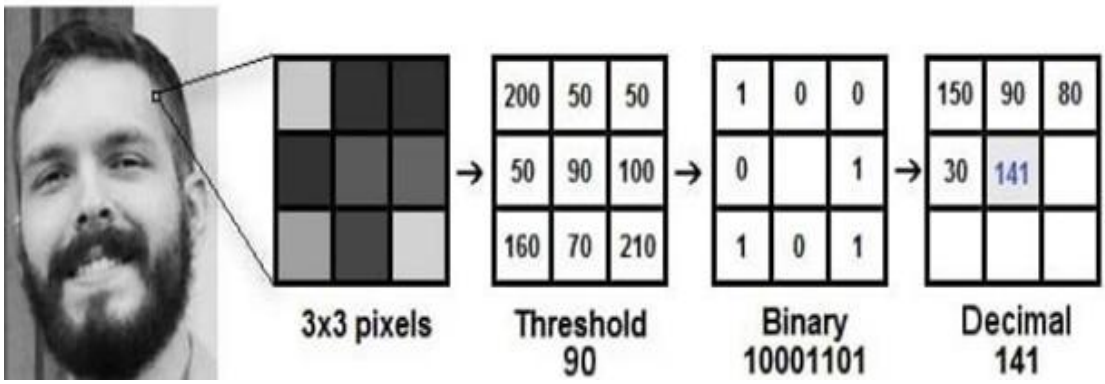

Figure 8: Local Binary Pattern Process

3. Trainer Algorithm

- **Step 1**: Initially we import the files which are required for training the machine i.e. CV, OS, NumPy and from Python Imaging Library (PIL) we want to import images.
- **Step 2**: For training the machine to recognize the face we create a face recognizer using face. LBPH Face Recognizer.
- **Step 3**: Creating the trail for dataset.
- **Step 4**: Create a detector by using Haar cascade.
- **Step 5**: We use function to urge the pictures and label
- **Step 6**: When the condition is completed the face sample and ids return.
- **Step 7**: Now the recognized faces and ids are sent to trainer for training.

- **Step 8**: Finally, the trained faces and ids are saved in trainer. Yml which is used fordetecting the face.

4. Detector Process

- Algorithms used in detector codes are designed to find frontal human faces.
- It is comparable to image detection, where an individual's image is matched bit by bit.
- •Images are compatible with those kept in databases.
- The database's matching process will be rendered useless if any face feature changes occur.
- Every potential candidate for a face is normalized to reduce both the lighting effect—caused by uneven illumination—and the shirring effect—due to head movement.

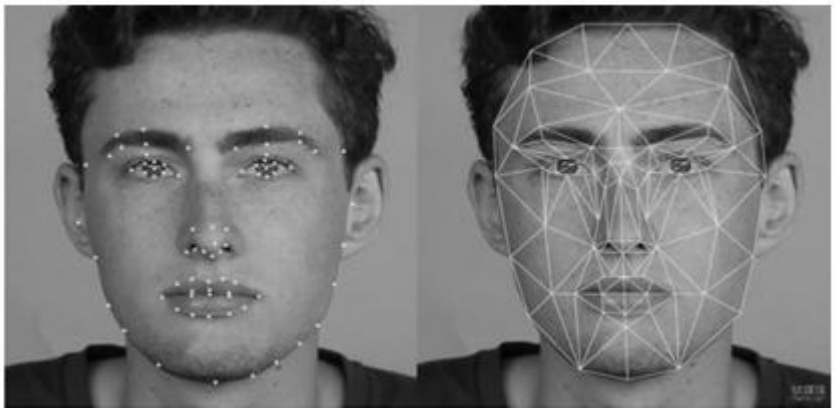

Figure 4: Face Detection Features

VII. RESULTS

Figure 5: Output Image

The above Figure represents the final output of the face recognition project. In this figure it represents the person who is already present in the data set and when the person is in-front of the camera, it recognizes and displays the name of the person on the display screen and respondswith a voice message by using pyaudio.

VIII. CONCLUSION

We developed a paradigm to evaluate the encounter discovery and acknowledgment procedures, which are regarded as a seat stamp, in the current work. A total of twenty-one data samples have been used to evaluate strategy performance since certain strategies operated consistently across several datasets while others continued to operate in an extremely erratic manner. The summary of the result of the confrontation location and acknowledgment method is provided in Fig 3. In the current framework, Haar-like [13] highlights are generally well-detailed, but they have significantly more incorrect discoveries than LBP [14], which might be taken into consideration as a future work in reconnaissance to reduce false locations in Haar-like [15] highlights. In contrast, Gabor [16] is well-detailed for the acknowledgment portion because its qualities outweigh the complexities.

IX. REFERENCES

[1] R. Jafri, H. R. Arabnia, “A Survey of Face Recognition Techniques”, Journal of Information Processing Systems, Vol.5, No.2, June 2009.
[2] C. A. Hansen, “Face Recognition”, Institute for Computer Science University of Tromso, Norway.
[3] M. D. Kelly. Visual identification of people by computer. PhD thesis, Stanford University,Stanford, CA, USA, 1971.
[4] T. Kanade. Computer Recognition of Human Faces, 47, 1977.
[5] W. Zhao, R. Chellappa, P. J. Phillips & A. Rosenfeld, “Face recognitions literature survey”, ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 399–458.
[6] C. Gonzalez, R. E. Woods, S. liddins, "Digital Image processing Using MATLAB".
[7] S. Suhas, A. Kurhe, Dr.P. Khanale, “Face Recognition Using Principal Component Analysis and Linear Discriminant Analysis on Holistic Approach in Facial Images Database”, IOSR Journal of Engineering e-ISSN: 2250-3021, p-ISSN: 2278-8719, Vol. 2, Issue 12 (Dec. 2012), ||V4|| PP 15-23
[8] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces", 1991.
[9] S. Asadi, Dr. D. V. Subba R. V. Saikrishna, "A Comparative study of Face Recognition with PCA and Cross-Correlation Technique", IJCA(0975-8887), Volume 10- No.8, November 2010.
[10] E. A. Abusham, A. T. B. Jin, W. E. Kiong, "Face Recognition Based on Nonlinear Feature Approach", American Journal of Applied Sciences, 2008.
[11] A. Nigam, P. Gupta, "A New Distance Measure for Face Recognition System", 2009 FifthInternational Conference on Image and Graphics
[12] A. Wager, Frances Wickham Lee and John P. Glaser.Managing health care information systems: A practicalapproach for health care executives (I), Jossey-Bass: A Wiley imprint San Francisco, 2005.
[13] Rajeev Agarwal-Voice Browsing the Web for Information access: http://www.w3.org/Voice/1998/Workshop/RajeevAgarwal.html
[14] T. Mita, T. Kaneko, O. Hori, Joint Haar-like Features forFace Detection, “Proceeding of the International Conference on Computer Vision”, 1550- Tenth IEEE 5499/05 ©2005 IEEE.
[15] T. Ahonen, A. Hadid, M. Peitikainen, Face recognition withlocal binary patterns. “In Proc.of European Conference ofComputer Vision”, 2004.
[16] L. Wiskott, M. Fellous, N. Krger, and C. Malsburg, Facerecognition by elastic bunch graph matching, “IEEE Trans”,on PAMI, 19:775–779, 1997
[17] K Ramesh Chandra, Madhusudan Donga, Prudhvi Raj Budumuru, Reversible Data Hiding Using Secure Image Transformation Technique, Inventive Systems and Control: Proceedings of ICISC 2021, 657-668
[18] D. MadhuSudhan, T Sindhu, An Advanced Method of Fragile Watermarking Using Local Binary Pattern (LBP), International Journal of Computer Science And Technology, 2016/4, Pages: 212-215
[19] Madhusudan Donga, Asha Korada, Prudhvi Raj Budumuru, Removal of impulse noise from a noisy image by using adaptive median filter, Journal of Xi'an University of Architecture & Technology, 2020/1/5, Pages :1385-1392.

