BIOMETRIC_AUTHENTICATION_FOR_ATMs.docx

by S D Hrudhay

RAMAIAH
Institute of Technology

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Artificial Intelligence - CS53

Non-CIE Component Report

Semester 5

Submitted by

S D HRUDHAY	1MS20CS098
SAFWAN G A	1MS20CS099
SANSKAR R GONDKAR	1MS20CS107
VAIEBHAV CHETTRI	1MS20CS134

Under the guidance of

Dr. S. Rajarajeswari
Associate Professor, Dept. of CSE

RAMAIAH INSTITUTE OF TECHNOLOGY
(Autonomous Institute, Affiliated to VTU)
BANGALORE-560054
www.msrit.edu, December 2022

RAMAIAH
Institute of Technology

CERTIFICATE

This is to certify that the Project work carried out by **S D HRUDHAY (1MS20CS098)**, **SAFWAN G A (1MS20CS099)**, **SANSKAR R GONDKAR (1MS20CS107)**, **VAIEBHAV CHETTRI (1MS20CS134)** as a 20-mark component for the course Artificial Intelligence (CS53), V semester B.E, CSE during the academic year September 2022- December 2022 satisfies the academic requirements for awarding the marks.

Signature of the Faculty

RAMAIAH
Institute of Technology

EVALUATION COMPONENTS

SL.NO	DESCRIPTION	MARKS
1.	Project Report	05
2.	Project Demo	05
3.	Project PPT/Presentation	05
4.	Seminar Presentation Slides	05

TABLE OF CONTENTS

RAMAIAH
Institute of Technology

BIOMETRIC AUTHENTICATION FOR ATMs

1. INTRODUCTION:

This paper presents a system for automated teller machine (ATM) authentication using face and voice recognition. The proposed system utilizes facial recognition to identify the customer and voice recognition to confirm their identity. The system is comprised of a camera and microphone for capturing facial and voice features, respectively. A feature extraction module is then used to extract the facial and voice features from the captured video and audio files respectively. The facial recognition component of the system uses a convolutional neural network (CNN) to extract features from the face image and compare them against a database of stored facial features. The speaker recognition system is based on a Long Short-Term Memory (LSTM) neural network trained on a large dataset of speech samples from the users. The voice recognition component utilizes a speech recognition algorithm to compare the customer's voice with that of the stored template. The system then combines the results from these two components to determine if the customer is an authorized user of the ATM. The proposed system has been tested on a dataset of four ATM users and results indicate that the system is able to accurately identify authorized users with an accuracy of 99.75%.

2. PROBLEM STATEMENT:

As the use of ATMs has grown, so have the number of incidents of fraud involving ATMs. Fraudsters have used a variety of methods to defraud ATM users, including skimming devices that capture account information and PINs, malicious software that captures account information and PINs, and shoulder surfing, where they watch users enter their PINs.

3. OBJECTIVES:

- Provide a more secure and convenient way for customers to access their accounts in ATMs:
    - A system that should be able to work in real-time and be robust against various types of spoofing attacks.
    - Prevent fraud and reduce the chances of people being able to access someone else's account in ATMs.
- A system that can authenticate a person based on their face and voice.
- To reduce the need for customers to remember PIN codes.

4. LITERATURE REVIEW:

Title	Technique used	Result	Remarks
Paper Title: Accuracy of facial recognition softwares. **Author Name:** M. A. Lazarini,R. Rossi and K. Hirama. **Source**: EAD Endorsed transaction. **Year**:2022	HAAR Cascade Algorithm, viola-Jones Algorithm, Fisher face algorithm, Binary Local Pattern Algorithm	Researched and found that all Algorithm used gave varying accuracies and established the differences in all the algorithm used.	We have understood that there are various algorithm for facial recognition and are still developing. We also learnt what algorithm we could for facial recognition.
Paper Title: Artificial Intelligence in Banking and Finance sectors **Author Names**: A. Geetha Assistant Professor, Department of Commerce, Dr. M.G.R. Educational and Research **Source:** International Journal of Creative Research Thoughts **Year**: 2021	Deep Research on AI in banking sector	Concluded that Banking and Financial Services consumers have good awareness about Artificial Intelligence applications.	We understood about the various applications which were being used and it helped us further evolve our project, after careful introspection.
Paper Title: Speech Recognition by Machine **Author Names**: M.A.Anusuya, S.K.Katti . Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore, India	Dynamic Time Warping, Robust Speech Recognition, Noisy Speech Recognition	Provides a comprehensive survey of research on speech recognition and establishes the importance of Robust Speech Recognition.	Understood the various concepts such as isolated Words, Completed words Continuous and Spontaneous speech. Also, developed understanding of various factors to be taken into account for speech recognition.

Source: International Journal of Computer science and Information Technology **Year:**2009			
Paper Title: Speech Recognition **Author Names**: Manjutha M, Gracy J, Dr P Subashini, Dr M Krishnaveni **Source:** international Journal of Engineering trends and Applications (IJETA) **Year:** 2017	VQLBG Algorithm, Conjugate Algorithm, Levenberg-Merquardt Algorithm.	The recognition accuracy is calculated for connected and continuous words, overall recognition rate for connected and continuous speech.	Understood various speech recognition techniques.
Paper Title: Application of Artificial Intelligence in Indian Banking-Opportunities and Challenges. **Authors Name**: Prof. Mohammed Nawaz, Prof. Triveni. K, Prof. Bharathi. G. R **Source**: International Journal of Trend in Scientific Research and Development **Year**:2020	With data analytics, block chain and machine learning, banks are advancing their services and offerings. The technology itself is getting better and smarter day by day, allowing more and newer banks to adopt the AI for various applications.	The aspects of understanding Artificial intelligence and its application in Indian banking sector and also analyses the opportunities and challenges of its application.	Image and face recognition using real time camera images and deep learning can be used at ATM's to prevent and detect frauds and crimes.
Paper Title: Automatic Speech Recognition: Systematic Literature Review **Author names**: SADEEN ALHARBI , MUNA ALRAZGAN,	Applying the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol	Automatic Speech Recognition is affected by 1.Nature of the Speech. 2.Vocabulary Size.	It helps us to know about the problem will be faced and are solved. And also the problem which are yet to be resilved

ALANOUD ALRASHED and collegues **Source**: IEEE Access **Year**:2021		3.Spectral Bandwidth. 4.Number of Speakers.	
Paper Title: Face Recognition: A Literature Survey **Authors Name**: W. Zhao,R Chellappa , P. J. Phillips, A. Rosenfeld **Source**: ACM Computing Surveys **Year**:2003	Holistic matching methods. Feature-based (structural) matching methods	Effect of lighting changing,Psychophysics/ neuroscience issues relevant to face recognition. Recognition rates were 100% and 93.9%, using 10 and 2 kernel discriminant analysis vectors, respectively.	Recognition of Face from the video is little time consuming than from the still photo because of the low quality and size of photo
Paper Title: Development of voice recognition: Parallels with face recognition **Author Name**: V. A. Mann, R. Diamond, And S. Carey **Source**: Journal Of Experimental Child Psychology **Year**:1979	Several studies suggest that children under age 10 represent highly familiar faces as do adults. Minimal change in voice recognition capacity occurs after age 4.	Performance of recognition changed markedly with age, improving sharply between ages 6 and 10, with 10-year-olds approaching adult levels. After age 10 accuracy declined significantly but returned to the adult level by age 14.	We use the face recognition for authentication as the voice may change because of health issues.
Paper Title: Face Recognition System **Authors Name**: Shivam Singh and Prof. S. Graceline Jasmine **Source**: International Journal of Engineering Research & Technology (IJERT) **Year**:2019	Proposed model 1.Image capture 2. Face Detection and Facial Features: Use viola Jones and KLT Algorithm Extract the Region of interest in Rectangular Bounding Box. 3.Pre-Processing: Convert to gray scale, apply histogram	The performance of this method is compared with other existing face recognition methods and it is observed that better accuracy in recognition is achieved with the proposed method. Face Recognition using KLT algorithm and Fusion of PCA and recognition plays a vital role in a	Using the same proposed methodology with the the changes in the algorithm for face feature extraction ,we'll do the face recognition

	equalization and Resize to 100x100 4. Database Development 5. Post-Processing	wide range of applications.	
Paper Title: Face Detection in Extreme Conditions: A Machine-learning Approach **Authors Title**: Sameer Aqib Hashmi	The cascade face detector proposed by using viola and jones makes use of haar-like features and AdaBoost to teach cascaded classifier.Convolution al neural network, MTCNN, Face Detection, Haar Cascade.	For different datasets Algorithm Accuracy Viola-Jones: 74.38% Haar Cascade :94% MTCNN :99.95% For their own datasets Algirithm Accuracy Haar Cascade: 68.16% MTCNN :98% Viola-Jones: 61.81%	Basic principles for the face recognition and the algorithm which is useful in extreme conditions
Paper Title: Face detection and Recognition: A review **Authors Name:** Akanksha, Jashanpreet Kaur and prof.Harjeet Singh **Source:** International Conference on Advancements in Engineering & Technology **Year:**2018	Adding the image to the database 1. Get the image. 2. Get the Face Detector object. 3. Apply the Face Detector object to the image to extract the features of detected face. 4. Add the image to the database. Comparing the input image with the database of images 1. Get the image. 2. Get the Face Detector object 3. Apply the Face Detector object to image and extract the features. 4. Compare the image with the database.	They've used 2 faces and in database the there are only 2 faces registered. The first face is registered face and the software is recognized face correctly. Second is not registered on the database and software says that he was not registered.	Different types of approach for the face detection like Feature base, holistic, hybrid.

Paper Title: VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT **Authors Name:** Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad **Source:** Journal of Applied and Fundamental Sciences	The feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). MFCC algorithm(to store extracted feature in .mat file) and VQ technique	The various approaches available for developing a Voice Recognition System based on adapted feature extraction technique and the speech recognition approach for the particular language are compared in this paper. The main aim of project is to develop a system that will allow the computer to translate voice request and dictation into text using MFCC and VQ techniques. Feature extraction and feature matching will be done using Mel Frequency Cepstral Coefficients and Vector Quantization technique.	Different types of speaker module and speech utterence result of other research paper helped us to understand the needs of project.
Paper Title: Voice Recognition and Voice Navigation for Blind using GPS **Authors Name:** Manish Bansode, Shivani Jadhav, Anjali Kashyap **Source:** INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING **Year:**2015	Software will get the distance between the two locations by longitude and latitude of destination and source.Alnd also find the current speed of the user. As the user moves the distance between the two location changes, speed of the user changes and subsequently the user gets the instruction.	Initially many blind people prefer to not use electronic aids, and use only canes or guide dogs. This is because of relatively high costs and poor levels of user satisfaction associated with existing electronic systems. So we tried to develop a low cost and user friendly system for blind people. This project can be extended by incorporating a GSM module. We can interface this module to send messages to the near and dear ones of the Blind person regarding his/her current position.	Different types of devices which is useful for the voice recognition and make it less cost

Paper Title: Human voice recognition depends on language ability **Authors name:** Tyler K. Perrachione, Stephanie N. Del Tufo, and John D.E. Gabrieli **Source:**NIH Public Access **Year:**2011	The ability to compute the discrepancies between the incidental phonetics of a particular vocalisation and the abstract phonological representations of the words that vocalisation contains is necessary for voice recognition. The human potential for voice recognition is severely hampered when the linguistic abstractions of words are absent or inadequate.	Human_speaking listeners with normal reading ability were significantly more accurate identifying voices speaking English than Chinese (paired t-test, P < 0.0005), performing on average 42% better in their native language (6). English-speaking listeners with dyslexia were no better able to identify English-speaking voices than Chinese-speaking ones (paired t-test, P = 0.65), with an average performance gain of only 2% in their native language	We need to consider how the speakers speech features according to that we need to recognize the speech

5. SOFTWARE REQUIREMENT SPECIFICATION:

5.1. Hardware required:

- Camera or Webcam to capture photo.
- Microphone for voice input.
- ARM or x86 processor with 2GHz or more.
- At least 2GB of free RAM to run the code with large dataset.
- 50GB of disk storage to store the data.
- Microsoft .NET 4.5 or higher for .NET component usage.
- Database to store user information.

5.2. Software required:

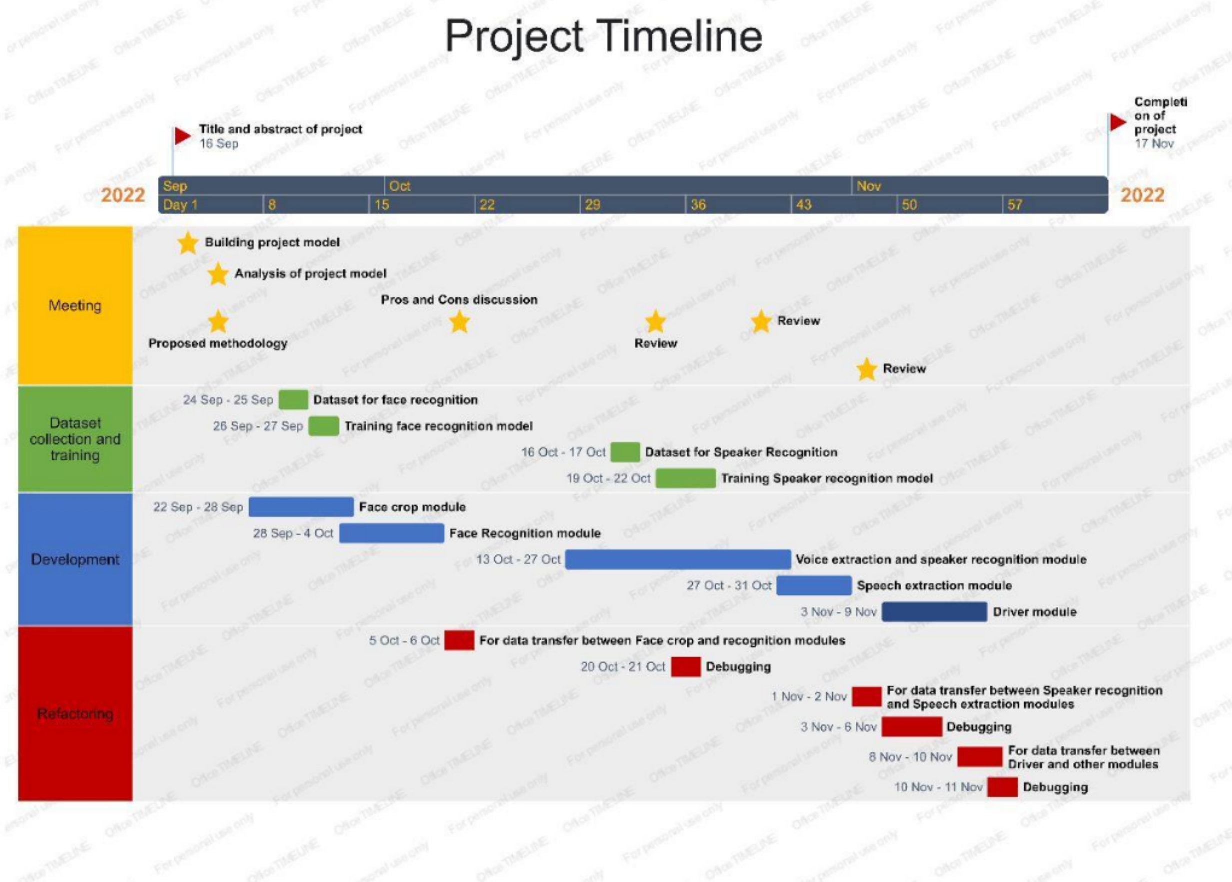
- Speech recognition system like GoogleAPI(online) or CMUShinx(offline) for speech-to-text conversion.
- Microsoft Visual Studio for application development under python environment.
- IBM RATIONAL ROSE SUITE for Uml diagrams.
- Kiwi to develop a mobile application interface.
- Google colab to train our CNN-model with sufficient VRAM.

RAMAIAH Institute of Technology

5.3. Budget:

Categories	Description	Cost (in rupees)
Hardware and Software procurement	No external hardware or software is needed to be procured at a cost as there are websites and software that provide resources for free for students.	0
Training AI - models	The face and voice recognition models were trained on Google Colab at free of cost.	0
Research	The research papers related to our project application for reference, were provided for free of cost by the authors and publishers.	0
Resources and Services	The Wi-Fi connection, research space and other needs were provided by the college at no cost. Dataset was for face and voice recognition models were found from various sources at no cost.	0
Total project cost (in rupees):		**0**

5.4. Timeline (Gantt chart):


Project Timeline
Title and abstract of project
16 Sep
Completi on of project
17 Nov
2022
Sep
Oct
Nov
Day 1
8
15
22
29
36
43
50
57
2022
Meeting
Building project model
Analysis of project model
Pros and Cons discussion
Proposed methodology
Review
Review
Review
Dataset collection and training
24 Sep - 25 Sep
Dataset for face recognition
26 Sep - 27 Sep
Training face recognition model
16 Oct - 17 Oct
Dataset for Speaker Recognition
19 Oct - 22 Oct
Training Speaker recognition model
Development
22 Sep - 28 Sep
Face crop module
28 Sep - 4 Oct
Face Recognition module
13 Oct - 27 Oct
Voice extraction and speaker recognition module
27 Oct - 31 Oct
Speech extraction module
3 Nov - 9 Nov
Driver module
Refactoring
5 Oct - 6 Oct
For data transfer between Face crop and recognition modules
20 Oct - 21 Oct
Debugging
1 Nov - 2 Nov
For data transfer between Speaker recognition and Speech extraction modules
3 Nov - 6 Nov
Debugging
8 Nov - 10 Nov
For data transfer between Driver and other modules
10 Nov - 11 Nov
Debugging

6. SYSTEM DESIGN:

6.1 Software Process Model:

Agile methodology is prevalent in situations where a minimal product is easy to make, and is expanded upon to make the final product in increments. Each of these increments is the result of a 'sprint', the cycle software engineers go through from conception of ideas all the way to implementing and testing it. Since the core software is relatively basic, it acts as a framework on which all other useful features are based on. This provides simplicity, since the developers will be focused on only a few things at a time, and reliability, since every new feature is tested to work with all old features when it is being implemented.

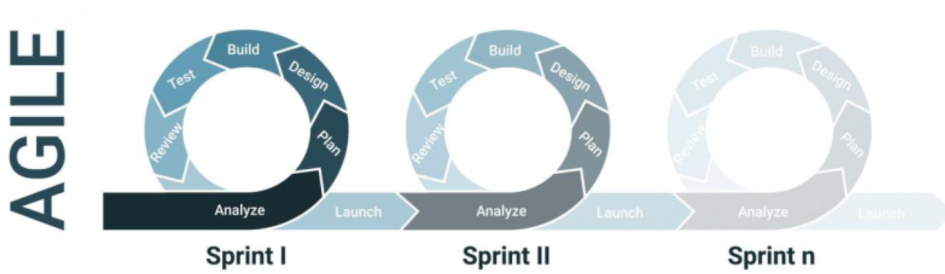


Figure 6.1: Illustration of Agile model.

The reason Agile development process was chosen for our project is that the core of the project is relatively easy to make, by interacting with a terminal and a small data set. It only needs to identify a few known people against a lot of others. Over time features need to incorporated that'll improve the usability of the software, like addition of more people faster, retraining faster and a GUI, all of which are not necessary to get it working at first. All of these need to be done in ordered sprints, since for example, we can't test the GUI for retraining feature if we haven't already added the feature before.

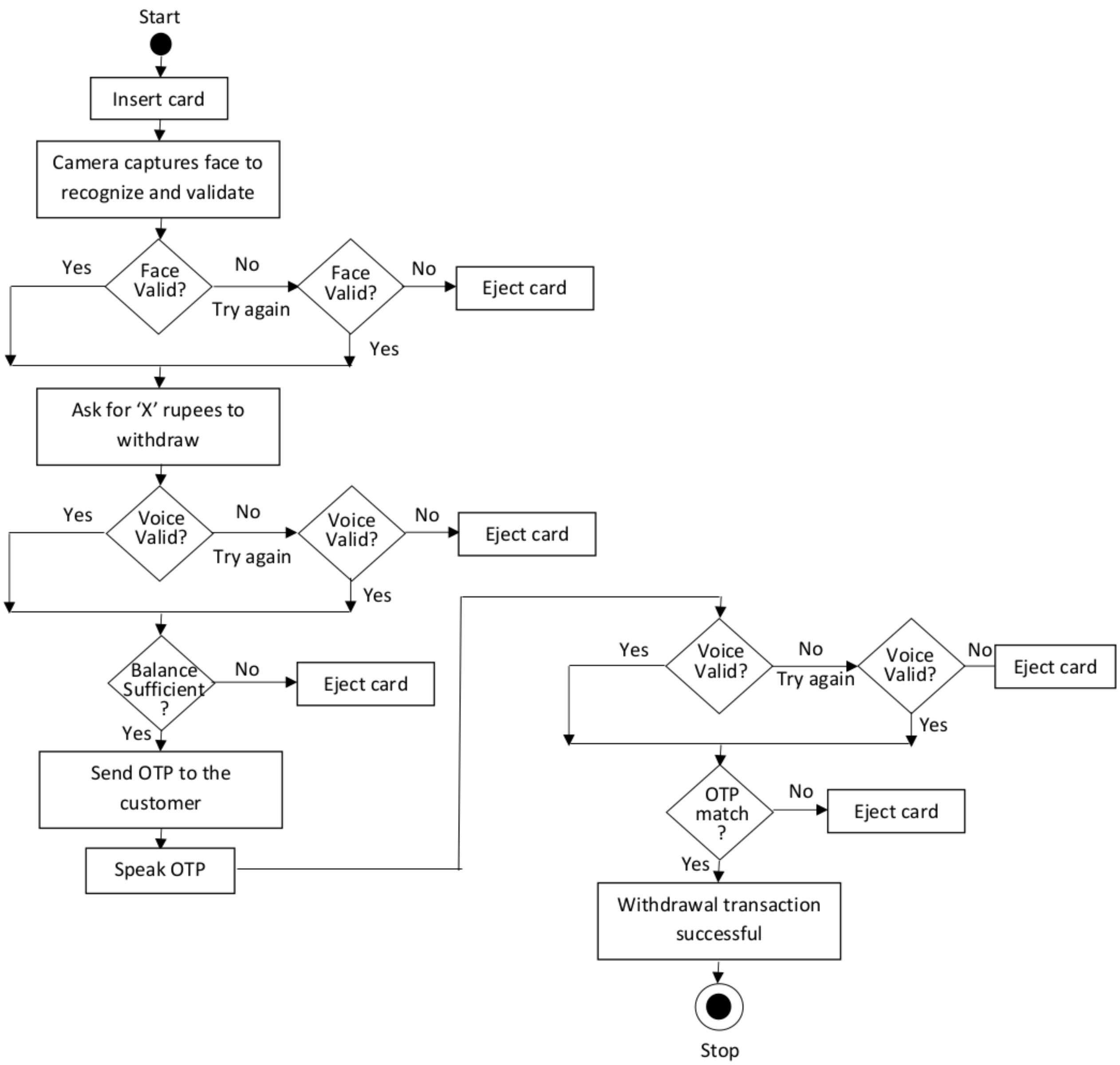


Figure 6.2: Activity diagram depicting the transaction marks.

Figure 6.3: Scenario based use case diagram.

Actors:

1.**User:** The user will insert the card, tell the machine how much money he wants, and then enter his OTP afterward. After this, he'll get paid

2.Atm machine: The ATM will simultaneously read the data from the card and take a picture, which it will subsequently send to the bank server. Furthermore, it accepts voice input for amount. It asks for voice input of the OTP after obtaining the notification that face recognition, voice matching and OTP transfer were successful. Once the OTP amount has been successfully verified, the user will be provided with amount.

3.Bank server: After obtaining the card information from the ATM, it uses the user id to query the database for the person's face and vocal features. It corresponds with the voice and facial input the ATM machine received. If a match is made, it then sends successful; otherwise, failure. After a successful verification, it sends an OTP to the registered mobile number, which it then recognizes and validates after hearing the OTP voice input. If successful, signals the ATM to release the funds, and the funds are then taken out of the account. If not, the transaction will fail.

Through the ATM machines, the user (bank customer) and the bank server are connected in this project. The ATM will read the card's information when the user inserts the card. It takes the picture while simultaneously checking it against database data on facial features.

The user is then prompted for the amount to be withdrawn before the withdrawal process is started. Additionally, following the elimination of the noise, it concurrently recognizes the speech and confirms the voice characteristic that was received.

The OTP will be issued to the registered mobile number if voice verification is completed. The money will be taken out of your bank account once your voice input for the OTP has been recognized and successfully verified.

The interaction between various actors with the specifier attributes is depicted in the sequence diagram as shown below:

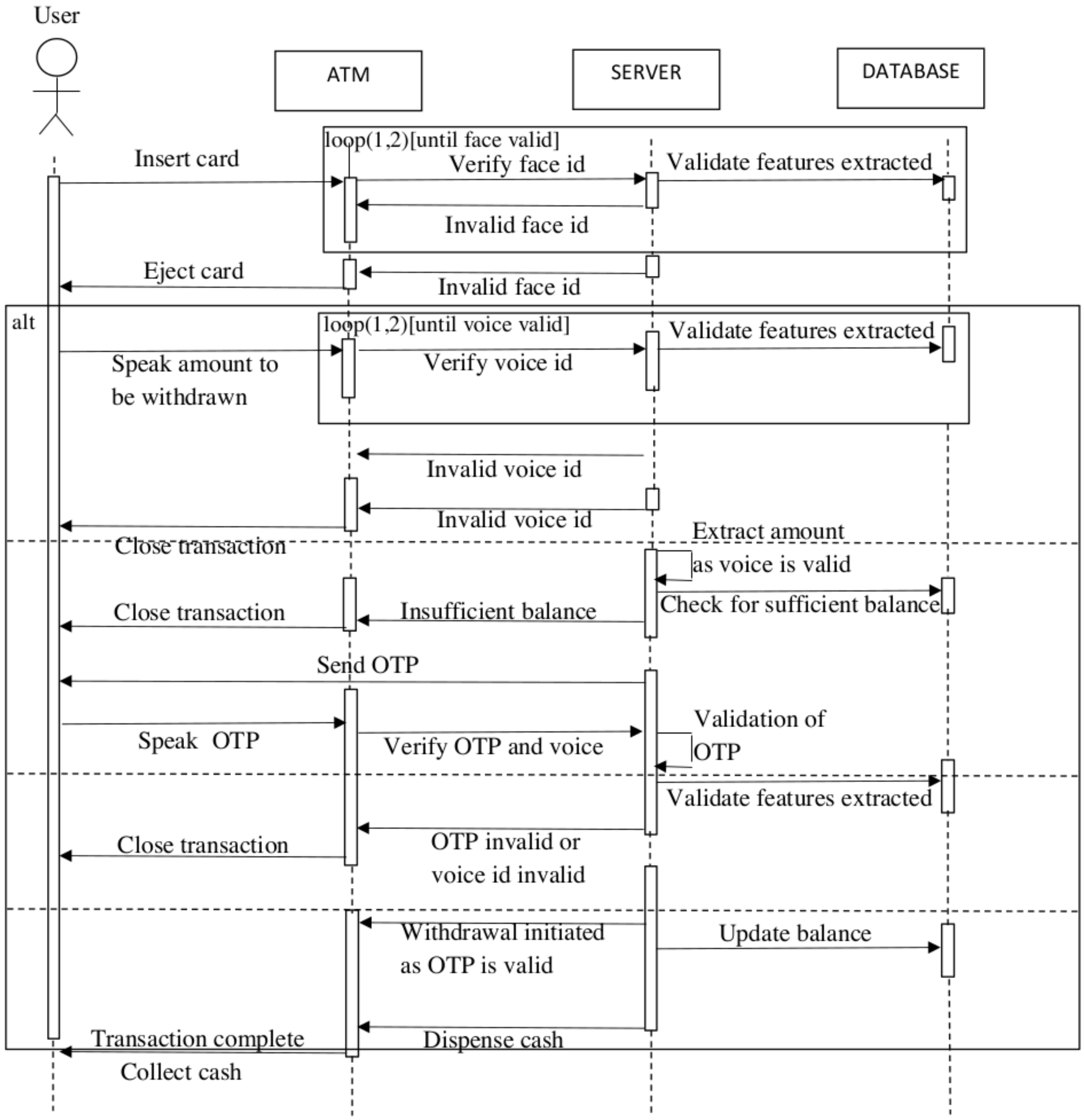


Figure 6.4: Sequence diagram for ATM system with AI enabled authentication.

7. <u>METHODOLOGY</u>:

7.1. Technology:

7.1.1. Face recognition:

The first step in face recognition is face detection, and this is done via systems such as Haar cascade.. The output of the Haar based program, which are all ideally of the same resolution, is the photos on which the face recognition program is run and contains the face to be recognized.

The kind of network we used to make our CNN model, is a Siamese network. Siamese networks, unlike conventional CNNs, don't classify the images into certain categories or labels, rather it only finds the likelihood of each pair of images given to it being the same person, i.e., label

A sigmoid Function is applied to the distance value to bring it to a range of 0-1. A loss function is applied on the sigmoid result, and that is used to update the weights and biases. Since both networks are a single instance, the updates done affect both equally.

Every-time convolution is applied, ReLu activation is used. This is to make sure that the values don't over correct themselves.

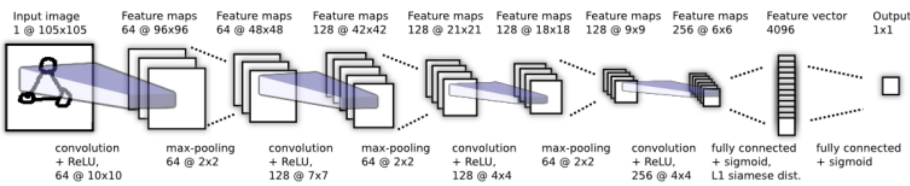


Figure 7.1.1.1 : Layered CNN Model for Face Recognition

7.1.2. Voice recognition:

Voice detection is done by audio classification software which can differentiate between human voices and everything else. This is relatively simple, and it consists of pre-processing the file so that all files are of same type and length, and then converting it to a spectrogram. The spectrogram is then passed through 2 convolution layers, then flattened and passed on to 2 dense layers. The output of the final layer is given sigmoid activation and all others ReLu. The final output determines if the voice is human or not.

Voice recognition model on the other hand, act on the same data, but only activates after a human voice is detected in an audio sample. The same audio sample is passed through this model. The idea is to make the correct identification of the speaker by using the Gaussian mixture model. The first step while dealing with an audio sample is to extract the features from it i.e. to identify components from the audio signal. We are using the Mel frequency cepstral coefficient (MFCC) to extract the features from the audio sample. MFCC which maps the signal onto a non-linear Mel-Scale that mimics the human hearing and provides the MFCC feature vectors which individually describes the power spectral envelope of a single frame.

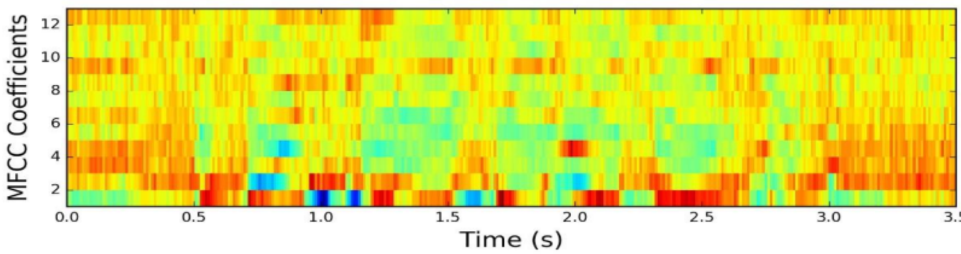


Figure 7.1.2.1: Spectrogram of MFCC coefficients.

We considered the MFCC with tuned parameter as a primary feature and delta MFCC which also known as differential and acceleration coefficients which are used to deal with speech information which is related to dynamics i.e., trajectories of MFCC coefficient over time it turns out to calculation of these trajectories. The model is trained using GMM. The GMM model will be used to calculate the scores of the features for all the models. The speaker model with the maximum score is predicted as the identified speaker of the test speech.

After voice is recognized to be the correct one, a few audio clips are combined and sent to speech recognition to find its contents. This is simple and is done with a speech to text conversion algorithm and thus won't be explained here in detail.

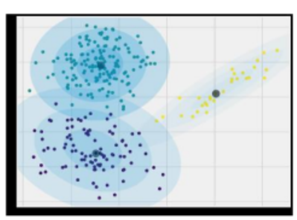


Figure 7.1.2.2: Illustration of GMM

7.2. Dataset:

The dataset used for training our face recognition model:

Labeled Faces in the Wild: A database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set. There are now four different sets of LFW images Among these, LFW-a and the deep funneled images produce superior results for most face verification algorithms over the original image.The facial images of the team members are also used for training the model.

The dataset used for training our speaker(voice) recognition model :

16000 pcm speeches: This dataset contains speeches of five prominent leaders namely; Benjamin Netanyahu, Jens Stoltenberg, Julia Gillard, Margaret Tacher and Nelson Mandela which also represents the folder names. Each audio in the folder is a one-second 16000 sample rate PCM encoded. The also contains voice samples of the team members.

7.3. Implementation:

7.3.1. Algorithm:

function AUTHENTICATE-USER(card) **returns** an action

state: card, currently inserted card

user_id, card data formulation

user_id ← EXTRACT-USER_ID(card)

if !AUTHENTICATE-FACE(user_id) **then**

return false

if !AUTHENTICATE_AMOUNT(user_id) **then**

return false

if !AUTHENTICATE_OTP(user_id) **then**

return false

else then

return true

RAMAIAH
Institute of Technology

function AUTHENTICATE-FACE(user_id) **returns** an action

state: face, capturing face percept

feature, face feature formulation

face ← CAPTURE-USER-FACE()

feature ← EXTRACT-FACIAL-FEATURE(face)

if FACIAL-MATCHING(face, user_id) **then**

return true

//if face authentication invalidates the user, he/she will be given a chance to re-validate

else if FACIAL-MATCHING(face, user_id) **then**

return true

else then

return false

function AUTHENTICATE_AMOUNT(user_id) **returns** an action

state: voice, audio input formulation

feature, vocal feature formulation

amount, amount formulation

voice ← ASK-FOR-AMOUNT()

feature ← EXTRACT-VOCAL-FEATURE(voice)

amount ← SPEECH-TO-TEXT(voice)

if VOCAL-MATCHING(feature, user_id) and SUFFICIENT-BALANCE(amount, user_id) **then**

return true

//if voice authentication invalidates the user, he/she will be given a chance to re-validate

else if VOCAL-MATCHING(feature, user_id) and SUFFICIENT-BALANCE(amount, user_id) **then**

return true

else then

return false

function AUTHENTICATE_OTP(user_id) **returns** an action

state: voice, audio input formulation

feature, vocal feature formulation

OTP, generated OTP

otp, extracted from the voice

OTP ← GENERATE-OTP()

voice ← ASK-FOR-OTP()

feature ← EXTRACT-VOCAL-FEATURE(voice)

otp ← SPEECH-TO-TEXT(voice)

if VOCAL-MATCHING(feature, user_id) and OTP-MATCHING(otp, OTP) **then**

return true

else then

return false

7.3.2. Pseudocode:

audiofromvideo.py

define path

take the video input

extract the voice and store it in the path

atmface.py

define the path

take frames from input video and store it

take some random frame and match with the data and count the hit

photofromvideo.py

define the path

input of the video

while there is a frame in video:

extract the photo and store it in the path

increment the counter

face_cropper.py

define path which is exist

detect the photo from the frame

crop the face and store it in the path

speech_extraction.py

define the path

recognize the voice in string type

convert to the int format and check if it matches with the otp for otp recognition

speaker_recognition.py

define the path which is exist

remove the noise from the voice and store it

extract the feature of the voice

match with the data

7.4. Testing:

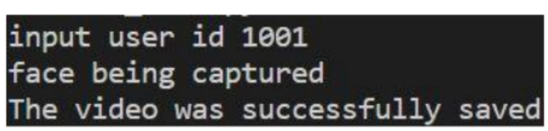


Figure 7.4.1: The beginning of the process and capturing the image.

When a user inserts a card into an ATM, a card reader included inside the machine scans the user ID or card number, and a video is also recorded for facial verification.

```
1/1 [==============================] - 0s 15ms/step
face verification failed
```

Figure .4.2: Test cases for the invalid user.

When a user attempts to use another person's ATM card, the software invalidates the user and cancels the transaction.

```
1/1 [==============================] - 0s 16ms/step
face verified suceesfully
say amount
Recording
Finished recording
amount is
5000
is input amount correct? (Y or N)Y
```

Figure 7.4.3: Test cases for the successful recognition and verification of user's face.

Following successful face verification, the ATM shall ask the user for the amount to be withdrawn and will employ voice for speech verification and voice recognition. Following the conversion from speech to text, software examines the amount to be withdrawn. And asks the user whether the amount is accurate for validation.

```
is input OTP correct? (Y or N)Y
1/1 [==============================] - 0s 16ms/step
second chance
1/1 [==============================] - 0s 15ms/step
voice verification failed
wrong OTP
```

Figure 7.4.4: Test cases for the invalid user voice.

If another person speaks OTP after successful facial verification, the user is rendered invalid and the transaction is terminated.

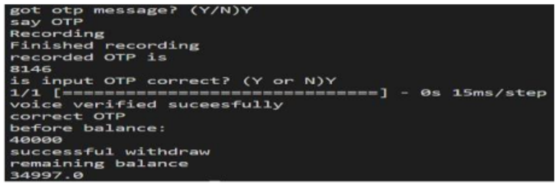


Figure 7.4.5: Test cases for the successful OTP verification and transaction.

Following amount verification, the software delivers the OTP to the registered mobile number. And, after the user obtains the OTP, it records the voice for the OTP. It verifies OTP by translating it to text format with the received OTP after user confirmation. And the amount will be transacted after successful OTP verification and user voice verification. It displays the amount before and after the transaction following the transaction.

7.5. Roles and Responsibility:

NAME (USN)	ROLES	RESPONSIBILITY
S D HRUDHAY (1MS20CS098)	Implementing programming, examining the project's general operation, and offering insightful information.	• Face Recognition. • Image to Data Conversion. • Training.
SAFWAN G A (1MS20CS099)	Evaluating and testing the project's efficient execution and giving it implementation dimensions.	• Face Detection. • Ignoring Background. • Process Video Frames. • Getting facial features contours.
SANSKAR GONDKAR (1MS20CS107)	The project's technical and programming components, as well as the recognition and use of the best operating procedures.	• Speaker Identification. • Open and closed set voice biometrics. • Training.
VAIEBHAV CHETTRI (1MS20CS134)	Supplying the project with the implementation dimensions and reviewing and testing the project's efficient execution.	• Speech Extraction. • Audio to Data Conversion. • Noise cancellation .

8. RESULTS AND DISCUSSIONS:

8.1. Performance metrics:

The performance metrics used for the efficiency of the project:

- Accuracy of the training and validation.
- Loss in the training and validation.
- Recall used for the how complete the model.

8.2. Results tabulation and graphs:

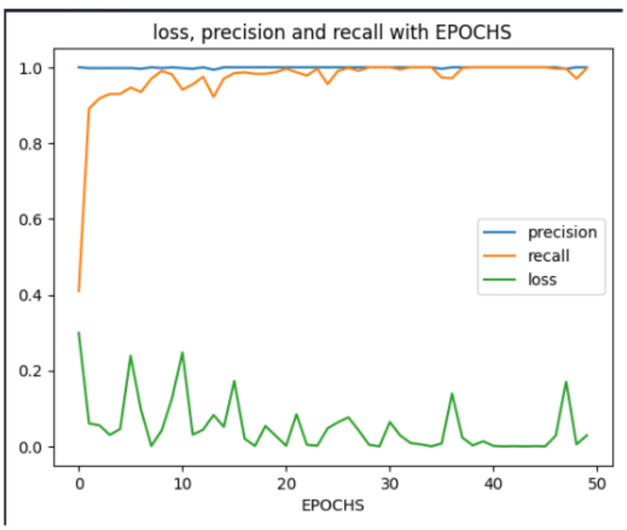


Figure 8.2.1: graph of the precision, loss, recall with respect to the epochs for the face.

Results obtained from the Figure 8.2.1 is shown below:

- Accuracy value nearly equal to 0.99.
- Recall value nearly equal to 0.99.
- Loss value nearly equal to 0.01.

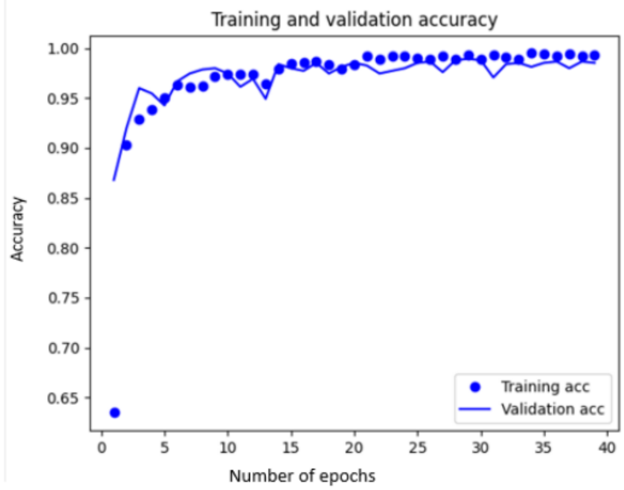


Figure 8.2.2: Accuracy of the model with respect to the number of epochs for the voice.

Results obtained from the Figure 8.2.2 is Accuracy value nearly equal to 0.99.

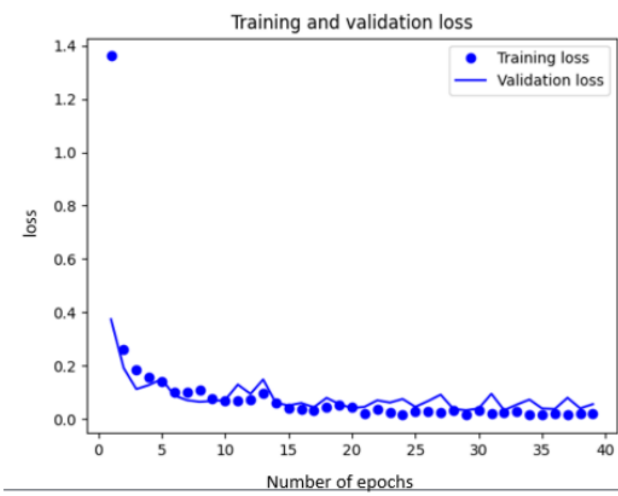


Figure 8.2.3: loss of the model with respect to the number of epochs for the voice.

Results obtained from the Figure 8.2.3 is loss value nearly equal to 0.01.

8.3. Comparison and analysis:

Figure 8.2.1 shows the face recognition model loss, accuracy, and recall with respect to the number of epochs. As we can see, recall—a metric of our model's completeness—increases in the vicinity of (0,8) epochs. After then, it rises linearly with little variation. Therefore, 8 epochs are the bare minimum needed for a satisfactory recall measure. Additionally, as the number of epochs rises, the precision and loss for face recognition increase and decrease respectively.

Figure 8.2.2 shows how our model's accuracy rises during training and validation in the range of (0,20) epochs. After then, the accuracy grows linearly with minimal variation and what appears to be a continual increase. Therefore, 20 epochs are the bare minimum needed for the good accuracy measure.

The loss in our model's training and validation decreases in the range of (0,20) epochs, as seen in Figure 8.2.3. After that, it decreases linearly with little variation and appears to be decreasing continuously without increasing accuracy. Therefore, 20 epochs are the bare minimum needed for the least loss measure.

So, 20 epochs are the bare minimum needed for successful recognition performance.

But we use 39 epochs for the higher the accuracy.

9. CONCLUSIONS:

In this project, the bank system's money withdrawals are authenticated using speech and face recognition. The project also generates OTP so that the user can benefit from a more secure service and without having to remember their pin number. The user's English accent will, however, affect the speech extraction. This offers the user a simple and safe means of making a payment.

10. SCOPE FOR THE FUTURE WORK:

Future work in this area could include

- AI-based authentication systems can be fooled by pictures or recordings of the person whose identity is being verified.
- AI-based systems can be slow and may not be able to handle large numbers of authentication requests in a short period of time i.e. they are difficult to implement and maintain.
- The accuracy of AI-based authentication systems can vary depending on the quality of the data used to train the system. Finally, AI-based authentication systems may not be compatible with all types of devices.
- The current model only supports English. To support the variety of local languages, the model needs to be modified.

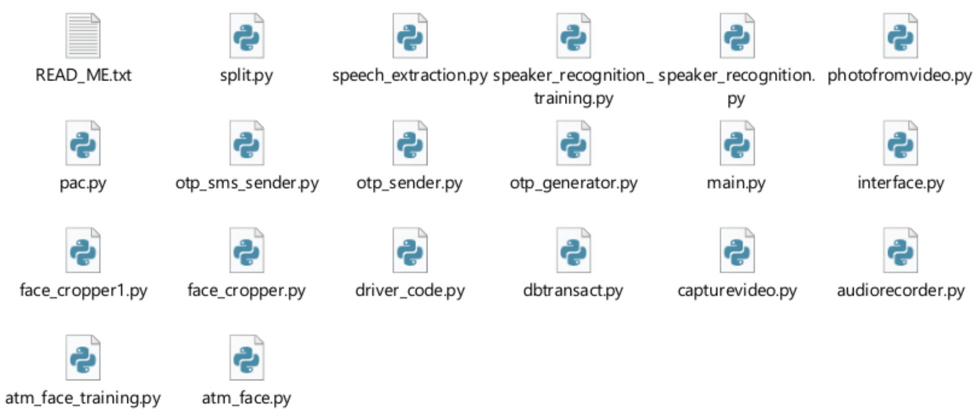
RAMAIAH
Institute of Technology

REFERENCE:

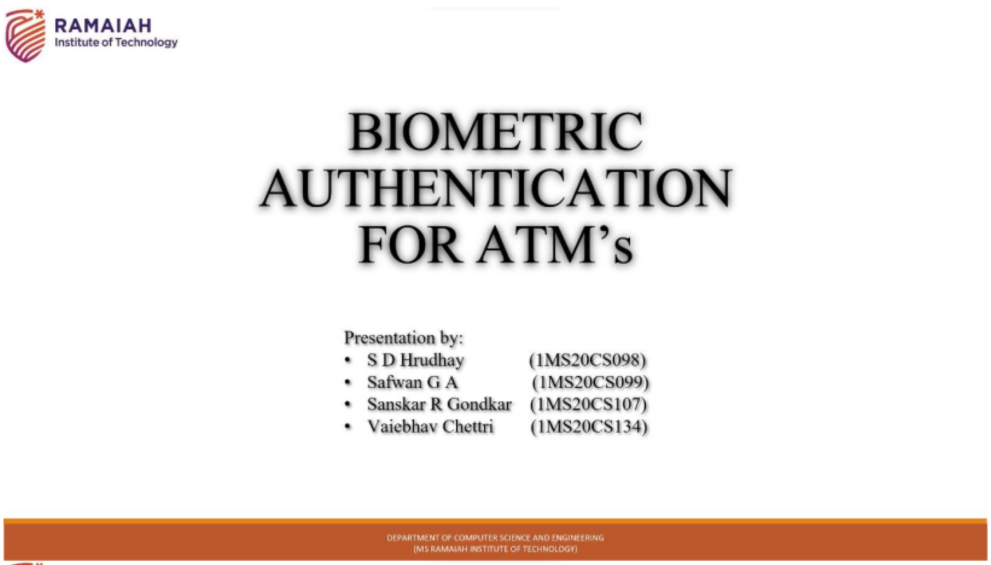
1. Practical Machine Learning and Image Processing_ For Facial Recognition, Object Detection, and Pattern Recognition Using Python by Himanshu Singh, APRESS publication 2019.
2. Automatic Speech Recognition_ A Deep Learning Approach by Dong Yu and Li Deng, Springer, 2014.
3. Face Detection and Recognition_ Theory and Practice by Asit Kumar, Madhura , Pradipta Kumar Banerjee, CRC PRESS, 2016.
4. Pro Deep Learning with TensorFlow by Santanu Pattanayak,APRESS publication, 2017.
5. 'Impact of Artificial Intelligence in chosen Indian commercial Bank-A Cost benefit analysis' by Sindhu J and Renee Namratha, Asiam journal of Management, ISSN 2321-5763, Volume 10, issue 4, 2019.
6. 'Artificial Intelligence in Banking-Mini Review' by Aoron Smith and Haitham Nobanee, SSRN electronic journal, Feb 2020.
7. Benton, A. L., & Van Allen, M. W. Manual: Test of fucial recognition. Nemosensory Center Publication No. 287, Department of Neurology, University Hospitals, Iowa City, IO, 1973.
8. PANTIC, M. AND ROTHKRANTZ, L. J. M. 2000. Automatic analysis of facial expressions: The state of the art. IEEE Trans. Patt. Anal. Mach. Intell. 22, 1424–1446.
9. O'TOOLE, A. J., ROARK, D., AND ABDI, H. 2002. Recognitizing moving faces. A psychological and neural synthesis. Trends Cogn. Sci. 6.
10. Accuracy of facial recognition softwares, M. A. Lazarini,R. Rossi and K. Hirama,EAD Endorsed transaction, 2022.
11. Artificial Intelligence in Banking and Finance sectors,A. Geetha Assistant Professor, Department of Commerce, Dr. M.G.R. Educational and Research, International Journal of Creative Research Thoughts, 2021.
12. Speech Recognition by Machine M.A.Anusuya, S.K.Katti . Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore, India,International Journal of Computer science and Information Technology, 2009.
13. Speech Recognition,Manjutha M, Gracy J, Dr P Subashini, Dr M Krishnaveni,international Journal of Engineering trends and Applications (IJETA), 2017.
14. Application of Artificial Intelligence in Indian Banking-Opportunities and Challenges,Prof. Mohammed Nawaz, Prof. Triveni. K, Prof. Bharathi. G. R ,International Journal of Trend in Scientific Research and Development, 2020.
15. Automatic Speech Recognition: Systematic Literature Review, SADEEN ALHARBI , MUNA ALRAZGAN, ALANOUD ALRASHED and collegues ,IEEE Access, 2021.
16. Face Recognition: A Literature Survey,W. Zhao,R Chellappa , P. J. Phillips, A. Rosenfeld ,ACM Computing Surveys, 2003.
17. Development of voice recognition:Parallels with face recognition,V. A. Mann, R. Diamond, And S. Carey,Journal Of Experimental Child Psychology, 1979.

18. Face Recognition System, Shivam Singh and Prof. S. Graceline Jasmine,International Journal of Engineering Research & Technology (IJERT), 2019.
19. Face Detection in Extreme Conditions: A Machine-learning Approach Sameer Aqib Hashmi.
20. Face detection and Recognition: A review,Akanksha ,Jashanpreet Kaur and prof.Harjeet Singh ,International Conference on Advancements in Engineering & Technology, 2018.
21. VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT,Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad,Journal of Applied and Fundamental Sciences.
22. Voice Recognition and Voice Navigation for Blind using GPS, Manisha Bansode, Shivani Jadhav, Anjali Kashyap, INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING ,2015.
23. Human voice recognition depends on language ability Tyler K. Perrachione, Stephanie N. Del Tufo, and John D.E. Gabrieli, NIH Public Access, 2011.

CODE:



PROJECT PRESENTATION SLIDES:



RAMAIAH
Institute of Technology

Introduction

Our project presents a system for automated teller machine (ATM) authentication using face and voice recognition. The proposed system utilizes facial recognition to identify the customer and voice recognition to confirm their identity. The system is comprised of a camera and microphone for capturing facial and voice features, respectively. A feature extraction module is then used to extract the facial and voice features from the captured video and audio files respectively. The facial recognition component of the system uses a convolutional neural network (CNN) to extract features from the face image and compare them against a database of stored facial features. The speaker recognition system is based on a Long Short-Term Memory (LSTM) neural network trained on a large dataset of speech samples from the users. The voice recognition component utilizes a speech recognition algorithm to compare the customer's voice with that of the stored template. The system then combines the results from these two components to determine if the customer is an authorized user of the ATM. The proposed system has been tested on a dataset of four ATM users and results indicate that the system is able to accurately identify authorized users with an accuracy of 99.75%.

RAMAIAH
Institute of Technology

Problem statement

“As the use of ATMs has grown, so have the number of incidents of fraud involving ATMs. Fraudsters have used a variety of methods to defraud ATM users, including skimming devices that capture account information and PINs, malicious software that captures account information and PINs, and shoulder surfing, where they watch users enter their PINs.”

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

Objectives

- Provide a more secure and convenient way for customers to access their accounts in ATMs:
    - A system that should be able to work in real-time and be robust against various types of spoofing attacks.
    - Prevent fraud and reduce the chances of people being able to access someone else's account in ATMs.
- A system that can authenticate a person based on their face and voice.
- To reduce the need for customers to remember PIN codes.

RAMAIAH
Institute of Technology

Literature review

Title	Technique used	Result	Remarks
Paper Title: Accuracy of facial recognition softwares **Author Name:** M. A. Lazarini,R. Rossi and K. Hirama **Source:** EAD Endorsed transaction **Year:**2022	HAAR Cascade Algorithm,viola-Jones Algorithm,Fisherface algorithm,Binary Local Pattern Algorithm	Researched and found that all Algorithm used gave varying accuracies and established the differences in all the algorithm used.	We have understood that there are various algorithm for facial recognition and are still developing. We also learnt what algorithm we could for facial recognition.
Paper Title: Artificial Intelligence in Banking and Finance sectors **Author Names:** A. Geetha Assistant Professor, Department of Commerce, Dr. M.G.R. Educational and Research **Source:** International Journal of Creative Research Thoughts **Year:** 2021	Deep Research on AI in banking sector	Concluded that Banking and Financial Services consumers have good awareness about Artificial Intelligence applications.	We understood about the various applications which were being used and it helped us further evolve our project, after careful introspection.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

Paper Title: Speech Recognition by Machine **Author Names**: M.A.Anusuya, S.K.Katti . Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore, India **Source:** International Journal of Computer science and Information Technology **Year:**2009	Dynamic Time Warping, Robust Speech Recognition, Noisy Speech Recognition	Provides a comprehensive survey of research on speech recognition and establishes the importance of Robust Speech Recognition.	Understood the various concepts such as isolated Words, Completed words Continuous and Spontaneous speech. Also, developed understanding of various factors to be taken into account for speech recognition.
Paper Title: Speech Recognition **Author Names**: Manjutha M, Gracy J, Dr P Subashini, Dr M Krishnaveni **Source:** international Journal of Engineering trends and Applications (IJETA) **Year:** 2017	VQLBG Algorithm, Conjugate Algorithm, Levenberg-Merquardt Algorithm.	The recognition accuracy is calculated for connected and continuous words, overall recognition rate for connected and continuous speech.	Understood various speech recognition techniques.

RAMAIAH
Institute of Technology

Paper Title: Application of Artificial Intelligence in Indian Banking- Opportunities and Challenges. **Authors Name**: Prof. Mohammed Nawaz, Prof. Triveni. K, Prof. Bharathi. G. R **Source**: International Journal of Trend in Scientific Research and Development **Year**:2020	With data analytics, block chain and machine learning, banks are advancing their services and offerings. The technology itself is getting better and smarter day by day, allowing more and newer banks to adopt the AI for various applications.	The aspects of understanding Artificial intelligence and its application in Indian banking sector and also analyses the opportunities and challenges of its application.	Image and face recognition using real time camera images and deep learning can be used at ATM's to prevent and detect frauds and crimes.
Paper Title: Automatic Speech Recognition: Systematic Literature Review **Author names**: SADEEN ALHARBI , MUNA ALRAZGAN, ALANOUD ALRASHED and collegues **Source**: IEEE Access **Year**:2021	Applying the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol	Automatic Speech Recognition is affected by 1.Nature of the Speech. 2.Vocabulary Size. 3.Spectral Bandwidth. 4.Number of Speakers.	It helps us to know about the problem will be faced and are solved. And also the problem which are yet to be resilved

RAMAIAH
Institute of Technology

Paper Title: Face Recognition: A Literature Survey **Authors Name**: W. Zhao,R Chellappa , P. J. Phillips, A. Rosenfeld **Source**: ACM Computing Surveys **Year**:2003	Holistic matching methods. Feature-based (structural) matching methods	Effect of lighting changing,Psychophysics/neuroscience issues relevant to face recognition. Recognition rates were 100% and 93.9%, using 10 and 2 kernel discriminant analysis vectors, respectively.	Recognition of Face from the video is little time consuming than from the still photo because of the low quality and size of photo
Paper Title: Development of voice recognition: Parallels with face recognition **Author Name**: V. A. Mann, R. Diamond, And S. Carey **Source**: Journal Of Experimental Child Psychology **Year**:1979	Several studies suggest that children under age 10 represent highly familiar faces as do adults. Minimal change in voice recognition capacity occurs after age 4.	Performance of recognition changed markedly with age, improving sharply between ages 6 and 10, with 10-year-olds approaching adult levels. After age 10 accuracy declined significantly but returned to the adult level by age 14.	We use the face recognition for authentication as the voice may change because of health issues.

RAMAIAH
Institute of Technology

Paper Title: Face Recognition System **Authors Name**: Shivam Singh and Prof. S. Graceline Jasmine **Source**: International Journal of Engineering Research & Technology (IJERT) **Year**:2019	Proposed model 1.Image capture 2. Face Detection and Facial Features: Use viola Jones and KLT Algorithm Extract the Region of interest in Rectangular Bounding Box. 3.Pre-Processing: Convert to gray scale, apply histogram equalization and Resize to 100x100 4. Database Development 5. Post-Processing	The performance of this method is compared with other existing face recognition methods and it is observed that better accuracy in recognition is achieved with the proposed method. Face Recognition using KLT algorithm and Fusion of PCA and recognition plays a vital role in a wide range of applications.	Using the same proposed methodology with the the changes in the algorithm for face feature extraction ,we'll do the face recognition
Paper Title: Face Detection in Extreme Conditions: A Machine-learning Approach **Authors Title**: Sameer Aqib Hashmi	The cascade face detector proposed by using viola and jones makes use of haar-like features and AdaBoost to teach cascaded classifier.Convolutional neural network, MTCNN, Face Detection, Haar Cascade.	For different datasets Algorithm Accuracy Viola-Jones: 74.38% Haar Cascade :94% MTCNN :99.95% For their own datasets Algirithm Accuracy Haar Cascade: 68.16% MTCNN :98% Viola-Jones: 61.81%	Basic principles for the face recognition and the algorithm which is useful in extreme conditions

RAMAIAH
Institute of Technology

Paper Title: Face detection and Recognition: A review **Authors Name:** Akanksha, Jashanpreet Kaur and prof.Harjeet Singh **Source:** International Conference on Advancements in Engineering & Technology **Year:**2018	Adding the image to the database 1. Get the image. 2. Get the Face Detector object. 3. Apply the Face Detector object to the image to extract the features of detected face. 4. Add the image to the database. Comparing the input image with the database of images 1. Get the image. 2. Get the Face Detector object 3. Apply the Face Detector object to image and extract the features. 4. Compare the image with the database.	They've used 2 faces and in database the there are only 2 faces registered. The first face is registered face and the software is recognized face correctly. Second is not registered on the database and software says that he was not registered.	Different types of approach for the face detection like Feature base, holistic, hybrid.
Paper Title: VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT **Authors Name:** Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad **Source:** Journal of Applied and Fundamental Sciences	The feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). MFCC algorithm(to store extracted feature in .mat file) and VQ technique	The various approaches available for developing a Voice Recognition System based on adapted feature extraction technique and the speech recognition approach for the particular language are compared in this paper. The main aim of project is to develop a system that will allow the computer to translate voice request and dictation into text using MFCC and VQ techniques. Feature extraction and feature matching will be done using Mel Frequency Cepstral Coefficients and Vector Quantization technique.	Different types of speaker module and speech utterence result of other research paper helped us to understand the needs of project.

RAMAIAH
Institute of Technology

Paper Title: Voice Recognition and Voice Navigation for Blind using GPS **Authors Name:** Manish Bansode, Shivani Jadhav, Anjali Kashyap **Source:** INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING **Year:**2015	Software will get the distance between the two locations by longitude and latitude of destination and source.Alnd also find the current speed of the user. As the user moves the distance between the two location changes, speed of the user changes and subsequently the user gets the instruction.	Initially many blind people prefer to not use electronic aids, and use only canes or guide dogs. This is because of relatively high costs and poor levels of user satisfaction associated with existing electronic systems. So we tried to develop a low cost and user friendly system for blind people. This project can be extended by incorporating a GSM module. We can interface this module to send messages to the near and dear ones of the Blind person regarding his/her current position.	Different types of devices which is useful for the voice recognition and make it less cost

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

Paper Title: Human voice recognition depends on language ability **Authors name:** Tyler K. Perrachione, Stephanie N. Del Tufo, and John D.E. Gabrieli **Source:**NIH Public Access **Year:**2011	The ability to compute the discrepancies between the incidental phonetics of a particular vocalisation and the abstract phonological representations of the words that vocalisation contains is necessary for voice recognition. The human potential for voice recognition is severely hampered when the linguistic abstractions of words are absent or inadequate.	Human_speaking listeners with normal reading ability were significantly more accurate identifying voices speaking English than Chinese (paired t-test, $P < 0.0005$), performing on average 42% better in their native language (6). English-speaking listeners with dyslexia were no better able to identify English-speaking voices than Chinese-speaking ones (paired t-test, $P = 0.65$), with an average performance gain of only 2% in their native language	We need to consider how the speakers speech features according to that we need to recognize the speech

System Design

Software Process Model:

Agile methodology is prevalent in situations where a minimal product is easy to make, and is expanded upon to make the final product in increments. Each of these increments is the result of a 'sprint', the cycle software engineers go through from conception of ideas all the way to implementing and testing it. Since the core software is relatively basic, it acts as a framework on which all other useful features are based on. This provides simplicity, since the developers will be focused on only a few things at a time, and reliability, since every new feature is tested to work with all old features when it is being implemented.

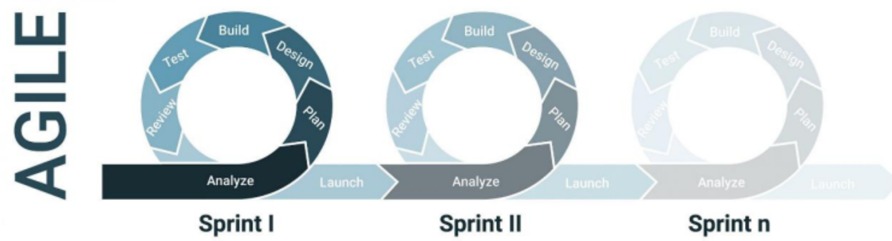


Figure 5.1: Software process model

The reason Agile development process was chosen for our project is that the core of the project is relatively easy to make, by interacting with a terminal and a small data set. It only needs to identify a few known people against a lot of others. Over time features need to incorporated that'll improve the usability of the software, like addition of more people faster, retraining faster and a GUI, all of which are not necessary to get it working at first. All of these need to be done in ordered sprints, since for example, we can't test the GUI for retraining feature if we haven't already added the feature before.

RAMAIAH
Institute of Technology

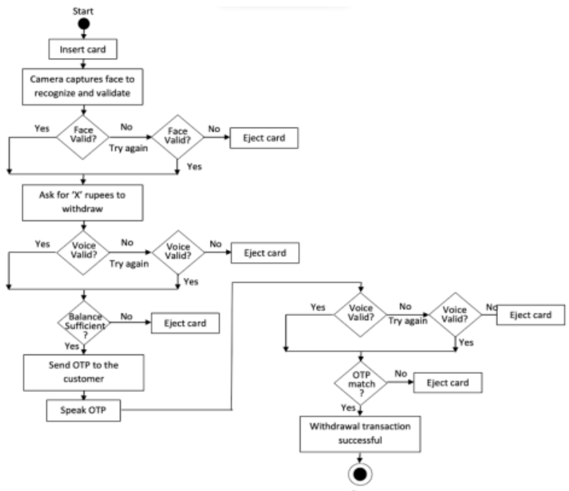


Figure 5.2: Activity diagram

RAMAIAH
Institute of Technology

Figure 5.3: Use-case diagram

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

Description of Actors:-

1.**USER:** The user will insert the card, tell the machine how much money he wants, and then enter his OTP afterward. After this, he'll get paid

2.ATM MACHINE: The ATM will simultaneously read the data from the card and take a picture, which it will subsequently send to the bank server. Furthermore, it accepts voice input for amount. It asks for voice input of the OTP after obtaining the notification that face recognition, voice matching and OTP transfer were successful. Once the OTP amount has been successfully verified, the user will be provided with amount.

3.BANK SERVER: After obtaining the card information from the ATM, it uses the user id to query the database for the person's face and vocal features. It corresponds with the voice and facial input the ATM machine received. If a match is made, it then sends successful; otherwise, failure. After a successful verification, it sends an OTP to the registered mobile number, which it then recognizes and validates after hearing the OTP voice input. If successful, signals the ATM to release the funds, and the funds are then taken out of the account. If not, the transaction will fail.

RAMAIAH
Institute of Technology

Through the ATM machines, the user (bank customer) and the bank server are connected in this project. The ATM will read the card's information when the user inserts the card. It takes the picture while simultaneously checking it against database data on facial features.

The user is then prompted for the amount to be withdrawn before the withdrawal process is started. Additionally, following the elimination of the noise, it concurrently recognizes the speech and confirms the voice characteristic that was received.

The OTP will be issued to the registered mobile number if voice verification is completed. The money will be taken out of your bank account once your voice input for the OTP has been recognized and successfully verified. The interaction between various actors with the specifier attributes is depicted in the sequence diagram as shown below:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH Institute of Technology

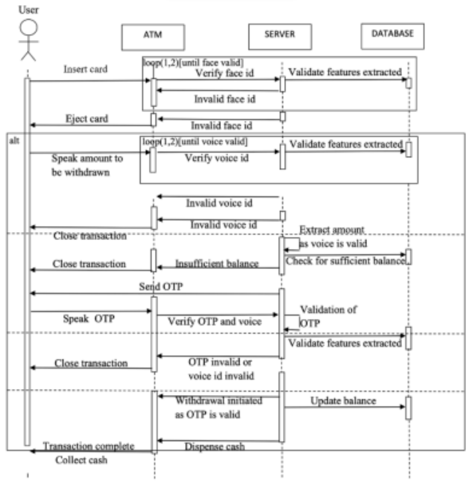


Figure 5.5: Sequence diagram

RAMAIAH Institute of Technology

Proposed Methodology

Technology:

Face recognition:

The first step in face recognition is face detection, and this is done via systems such as Haar cascade.. The output of the Haar based program, which are all ideally of the same resolution, is the photos on which the face recognition program is run and contains the face to be recognized.

The kind of network we used to make our CNN model, is a Siamese network. Siamese networks, unlike conventional CNNs, don't classify the images into certain categories or labels, rather it only finds the likelihood of each pair of images given to it being the same person, i.e., label

A sigmoid Function is applied to the distance value to bring it to a range of 0-1. A loss function is applied on the sigmoid result, and that is used to update the weights and biases. Since both networks are a single instance, the updates done affect both equally.

Every-time convolution is applied, ReLu activation is used. This is to make sure that the values don't over correct themselves.

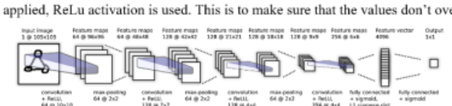


Figure 6.1.1: Illustration of the CNN model

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

Voice recognition:

Voice detection is done by audio classification software which can differentiate between human voices and everything else. This is relatively simple, and it consists of pre-processing the file so that all files are of same type and length, and then converting it to a spectrogram. The spectrogram is then passed through 2 convolution layers, then flattened and passed on to 2 dense layers. The output of the final layer is given sigmoid activation and all others ReLu. The final output determines if the voice is human or not.

Voice recognition model on the other hand, act on the same data, but only activates after a human voice is detected in an audio sample. The same audio sample is passed through this model. The idea is to make the correct identification of the speaker by using the Gaussian mixture model. The first step while dealing with an audio sample is to extract the features from it i.e. to identify components from the audio signal. We are using the Mel frequency cepstral coefficient (MFCC) to extract the features from the audio sample. MFCC which maps the signal onto a non-linear Mel-Scale that mimics the human hearing and provides the MFCC feature vectors which individually describes the power spectral envelope of a single frame.

We considered the MFCC with tuned parameter as a primary feature and delta MFCC which also known as differential and acceleration coefficients which are used to deal with speech information which is related to dynamics i.e., trajectories of MFCC coefficient over time it turns out to calculation of these trajectories. The model is trained using GMM. The GMM model will be used to calculate the scores of the features for all the models. The speaker model with the maximum score is predicted as the identified speaker of the test speech.

After voice is recognized to be the correct one, a few audio clips are combined and sent to speech recognition to find its contents. This is simple and is done with a speech to text conversion algorithm and thus won't be explained here in detail.

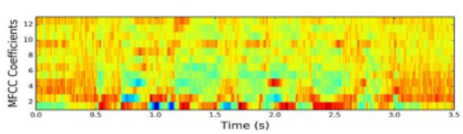


Figure 6.1.2: Spectrogram of MFCC coefficients

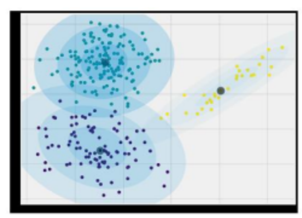


Figure 6.1.3: Illustration of GMM

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

Dataset:

The dataset used for training our face recognition model :

Labeled Faces in the Wild: A database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set. There are now four different sets of LFW images Among these, LFW-a and the deep funneled images produce superior results for most face verification algorithms over the original image.The facial images of the team members are also used for training the model.

The dataset used for training our speaker(voice) recognition model :

16000 pcm speeches: This dataset contains speeches of five prominent leaders namely; Benjamin Netanyahu, Jens Stoltenberg, Julia Gillard, Margaret Tacher and Nelson Mandela which also represents the folder names. Each audio in the folder is a one-second 16000 sample rate PCM encoded. The also contains voice samples of the team members.

Algorithm :

function AUTHENTICATE-USER(card) **returns** an action

state: card, currently inserted card

user_id, card data formulation

user_id ← EXTRACT-USER_ID(card)

if !AUTHENTICATE-FACE(user_id) **then**

return false

if !AUTHENTICATE_AMOUNT(user_id) **then**

return false

if !AUTHENTICATE_OTP(user_id) **then**

return false

else then

return true

RAMAIAH
Institute of Technology

```
function AUTHENTICATE-FACE(user_id) returns an action
state:  face, capturing face percept
        feature, face feature formulation
face    ←  CAPTURE-USER-FACE()
feature ←  EXTRACT-FACIAL-FEATURE(face)
if FACIAL-MATCHING(face, user_id) then
        return true
//if face authentication invalidates the user, he/she will be given a chance to re-validate
else if FACIAL-MATCHING(face, user_id) then
        return true
else then
        return false
```


RAMAIAH
Institute of Technology

```
function AUTHENTICATE_AMOUNT(user_id) returns an action
state:  voice, audio input formulation
        feature, vocal feature formulation
        amount, amount formulation
voice   ←  ASK-FOR-AMOUNT()
feature ←  EXTRACT-VOCAL-FEATURE(voice)
amount  ←  SPEECH-TO-TEXT(voice)
if VOCAL-MATCHING(feature, user_id) and SUFFICIENT-BALANCE( amount, user_id) then
        return true
//if voice authentication invalidates the user, he/she will be given a chance to re-validate
else if VOCAL-MATCHING(feature, user_id) and SUFFICIENT-BALANCE( amount, user_id) then
        return true
else then
        return false
```

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

function AUTHENTICATE_OTP(user_id) **returns** an action

state: voice, audio input formulation

feature, vocal feature formulation

OTP, generated OTP

otp, extracted from the voice

OTP ← GENERATE-OTP()

voice ← ASK-FOR-OTP()

feature ← EXTRACT-VOCAL-FEATURE(voice)

otp ← SPEECH-TO-TEXT(voice)

if VOCAL-MATCHING(feature, user_id) and OTP-MATCHING(otp, OTP) **then**

return true

else then

return false

Pseudocode:

audiofromvideo.py
- define path
- take the video input
- extract the voice and store it in the path

atmface.py
- define the path
- take frames from input video and store it
- take some random frame and match with the data and count the hit

photofromvideo.py
- define the path
- input of the video
- while there is a frame in video:
  - extract the photo and store it in the path
  - increment the counter

face_cropper.py
- define path which is exist
- detect the photo from the frame
- crop the face and store it in the path

RAMAIAH
Institute of Technology

speech_extraction.py

define the path

recognize the voice in string type

convert to the int format and check if it matches with the otp for otp recognition

speaker_recognition.py

define the path which is exist

remove the noise from the voice and store it

extract the feature of the voice

match with the data

RAMAIAH
Institute of Technology

Testing:

```
input user id 1001
face being captured
The video was successfully saved
```

Figure 6.4.1: The beginning of the process and capturing the image

When a user inserts a card into an ATM, a card reader included inside the machine scans the user ID or card number, and a video is also recorded for facial verification.

```
1/1 [==============================] - 0s 15ms/step
face verification failed
```

Figure 6.4.2: Test cases for the invalid user.

When a user attempts to use another person's ATM card, the software invalidates the user and cancels the transaction.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

```
1/1 [==============================] - 0s 16ms/step
face verified suceesfully
say amount
Recording
Finished recording
amount is
5000
is input amount correct? (Y or N)Y
```

Figure 6.4.3: Test cases for the successful recognition and verification of user's face.

Following successful face verification, the ATM shall ask the user for the amount to be withdrawn and will employ voice for speech verification and voice recognition. Following the conversion from speech to text, software examines the amount to be withdrawn. And asks the user whether the amount is accurate for validation.

```
is input OTP correct? (Y or N)Y
1/1 [==============================] - 0s 16ms/step
second chance
1/1 [==============================] - 0s 15ms/step
voice verification failed
wrong OTP
```

Figure 6.4.4: Test cases for the invalid user voice.

If another person speaks OTP after successful facial verification, the user is rendered invalid and the transaction is terminated.

RAMAIAH
Institute of Technology

```
got otp message? (Y/N)Y
say OTP
Recording
Finished recording
recorded OTP is
8146
is input OTP correct? (Y or N)Y
1/1 [==============================] - 0s 15ms/step
voice verified suceesfully
correct OTP
before balance:
40000
successful withdraw
remaining balance
34997.0
```

Figure 6.4.1: Test cases for the successful OTP verification and transaction.

Following amount verification, the software delivers the OTP to the registered mobile number. And, after the user obtains the OTP, it records the voice for the OTP. It verifies OTP by translating it to text format with the received OTP after user confirmation. And the amount will be transacted after successful OTP verification and user voice verification. It displays the amount before and after the transaction following the transaction.

Roles and Responsibility:

NAME (USN)	ROLES	RESPONSIBILITY
S D HRUDHAY (1MS20CS098)	Implementing programming, examining the project's general operation, and offering insightful information.	• Face Recognition. • Image to Data Conversion. • Training.
SAFWAN G A (1MS20CS099)	Evaluating and testing the project's efficient execution and giving it implementation dimensions.	• Face Detection. • Ignoring Background. • Process Video Frames. • Getting facial features contours.
SANSKAR GONDKAR (1MS20CS107)	The project's technical and programming components, as well as the recognition and use of the best operating procedures.	• Speaker Identification. • Open and closed set voice biometrics. • Training.
VAIEBHAV CHETTRI (1MS20CS134)	Supplying the project with the implementation dimensions and reviewing and testing the project's efficient execution.	• Speech Extraction. • Audio to Data Conversion. • Noise cancellation.

Results and Discussions

Performance metrics:

The performance metrics used for the efficiency of the project are

- Accuracy of the training and validation.
- Loss in the training and validation.
- Recall used for the how complete the model.

RAMAIAH
Institute of Technology

Results tabulation and graphs:

At the end, we got the precision and recall of face recognition value nearly 0.99 and the loss nearly 0.01.

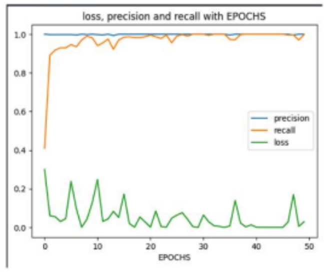


Figure 7.2.1: graph of the precision, loss, recall with respect to the epochs for the face.

At the end, we got the accuracy of voice recognition nearly equal to 0.99.

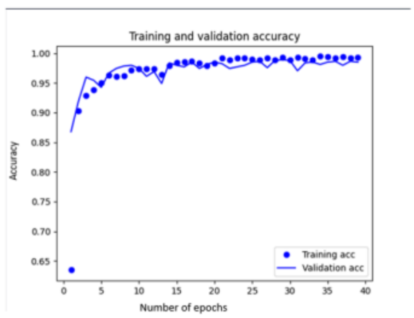


Figure 7.2.2: Accuracy of the model with respect to the number of epochs for the voice.

RAMAIAH
Institute of Technology

At the end, we got the loss of voice recognition nearly equal to 0.01.

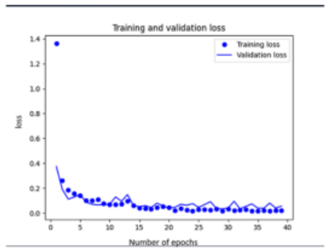


Figure 7.2.3: loss of the model with respect to the number of epochs for the voice.

RAMAIAH
Institute of Technology

Comparison and analysis:

Figure 7.2.1 shows the face recognition model loss, accuracy, and recall with respect to the number of epochs. As we can see, recall—a metric of our model's completeness—increases in the vicinity of (0,8) epochs. After then, it rises linearly with little variation. Therefore, 8 epochs are the bare minimum needed for a satisfactory recall measure. Additionally, as the number of epochs rises, the precision and loss for face recognition increase and decrease respectively.

Figure 7.2.2 shows how our model's accuracy rises during training and validation in the range of (0,20) epochs. After then, the accuracy grows linearly with minimal variation and what appears to be a continual increase. Therefore, 20 epochs are the bare minimum needed for the good accuracy measure.

The loss in our model's training and validation decreases in the range of (0,20) epochs, as seen in Fig. 7.2.3. After that, it decreases linearly with little variation and appears to be decreasing continuously without increasing accuracy. Therefore, 20 epochs are the bare minimum needed for the least loss measure.

So, 20 epochs are the bare minimum needed for successful recognition performance.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

But, we use 30 epochs for the higher the accuracy and for which the performance is given by,

For the face,

- Accuracy value nearly equal to 0.99
- Recall value nearly equal to 0.99
- Loss value nearly equal to 0.01

For the voice,

- Accuracy value nearly equal to 0.99
- Loss value nearly equal to 0.01

Conclusion

In this project, the bank system's money withdrawals are authenticated using speech and face recognition. The project also generates OTP so that the user can benefit from a more secure service and without having to remember their pin number. The user's English accent will, however, affect the speech extraction. This offers the user a simple and safe means of making a payment.

Scope for Future Work

Future scope of our project includes,

- AI-based authentication systems can be fooled by pictures or recordings of the person whose identity is being verified.
- AI-based systems can be slow and may not be able to handle large numbers of authentication requests in a short period of time i.e. they are difficult to implement and maintain.
- The accuracy of AI-based authentication systems can vary depending on the quality of the data used to train the system. Finally, AI-based authentication systems may not be compatible with all types of devices.
- Only English is supported by the present model. The model has to be improved to enable diverse languages that are spoken locally.

References

1. Practical Machine Learning and Image Processing_ For Facial Recognition, Object Detection, and Pattern Recognition Using Python by Himanshu Singh, APRESS publication 2019.
2. Automatic Speech Recognition_ A Deep Learning Approach by Dong Yu and Li Deng, Springer, 2014.
3. Face Detection and Recognition_ Theory and Practice by Asit Kumar, Madhura , Pradipta Kumar Banerjee, CRC PRESS, 2016.
4. Pro Deep Learning with TensorFlow by Santanu Pattanayak,APRESS publication, 2017.
5. 'Impact of Artificial Intelligence in chosen Indian commercial Bank-A Cost benefit analysis' by Sindhu J and Renee Namratha, Asiam journal of Management, ISSN 2321-5763, Volume 10, issue 4, 2019.
6. 'Artificial Intelligence in Banking-Mini Review' by Aoron Smith and Haitham Nobanee, SSRN electronic journal, Feb 2020.
7. Benton, A. L., & Van Allen, M. W. Manual: Test of fucial recognition. Nemosensory Center Publication No. 287, Department of Neurology, University Hospitals, Iowa City, IO, 1973.
8. PANTIC, M. AND ROTHKRANTZ, L. J. M. 2000. Automatic analysis of facial expressions: The state of the art. IEEE Trans. Patt. Anal. Mach. Intell. 22, 1424–1446.

RAMAIAH
Institute of Technology

RAMAIAH
Institute of Technology

9. Automatic Speech Recognition: Systematic Literature Review, SADEEN ALHARBI , MUNA ALRAZGAN, ALANOUD ALRASHED and collegues ,IEEE Access, 2021.
10. Face Recognition: A Literature Survey,W. Zhao,R Chellappa , P. J. Phillips, A. Rosenfeld ,ACM Computing Surveys, 2003.
11. Development of voice recognition:Parallels with face recognition,V. A. Mann, R. Diamond, And S. Carey,Journal Of Experimental Child Psychology, 1979.
12. Face Recognition System, Shivam Singh and Prof. S. Graceline Jasmine,International Journal of Engineering Research & Technology (IJERT), 2019.
13. Face Detection in Extreme Conditions: A Machine-learning Approach Sameer Aqib Hashmi.
14. Face detection and Recognition: A review,Akanksha ,Jashanpreet Kaur and prof.Harjeet Singh ,International Conference on Advancements in Engineering & Technology, 2018.
15. VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT,Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad,Journal of Applied and Fundamental Sciences.
16. Voice Recognition and Voice Navigation for Blind using GPS, Manisha Bansode, Shivani Jadhav, Anjali Kashyap, INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING ,2015.
17. Human voice recognition depends on language ability Tyler K. Perrachione, Stephanie N. Del Tufo, and John D.E. Gabrieli, NIH Public Access, 2011.

RAMAIAH
Institute of Technology

18. O'TOOLE, A. J., ROARK, D., AND ABDI, H. 2002. Recognitizing moving faces. A psychological and neural synthesis. Trends Cogn. Sci. 6.
19. Accuracy of facial recognition softwares, M. A. Lazarini,R. Rossi and K. Hirama,EAD Endorsed transaction, 2022.
20. Artificial Intelligence in Banking and Finance sectors,A. Geetha Assistant Professor, Department of Commerce, Dr. M.G.R. Educational and Research, International Journal of Creative Research Thoughts, 2021.
21. Speech Recognition by Machine M.A.Anusuya, S.K.Katti . Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore, India,International Journal of Computer science and Information Technology, 2009.
22. Speech Recognition,Manjutha M, Gracy J, Dr P Subashini, Dr M Krishnaveni,international Journal of Engineering trends and Applications (IJETA), 2017.
23. Application of Artificial Intelligence in Indian Banking-Opportunities and Challenges,Prof. Mohammed Nawaz, Prof. Triveni. K, Prof. Bharathi. G. R ,International Journal of Trend in Scientific Research and Development, 2020.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(MS RAMAIAH INSTITUTE OF TECHNOLOGY)

RAMAIAH
Institute of Technology

THANK YOU

RAMAIAH
Institute of Technology

SEMINAR PRESENTATION SLIDES:

RAMAIAH
Institute of Technology

ROBOTIC PERCEPTION

Presentation by:

- S D Hrudhay (1MS20CS098)
- Safwan G A (1MS20CS099)
- Sanskar R Gondkar (1MS20CS107)
- Vaiebhav Chettri (1MS20CS134)

RAMAIAH
Institute of Technology

Perception is the process by which robots map sensor measurements into internal representations of the environment.

Sometimes it is difficult for robots to percept.

In what way is perception difficult?

- Sensors are noisy.
- Environment is partially observable.
- Environment is unpredictable.
- Environment is dynamic.

Good internal representations for robots have three properties:

- They contain enough information for the robot to make good decisions.
- They are structured so that they can be updated efficiently.
- They are natural in the sense that internal variables correspond to natural state variables in the physical world.

Due to uncertainty in perception as discussed above it becomes difficult for the robot to estimate the state. Instead, the state of the robot is best characterized by a probability distribution function (pdf) over all possible states, which is commonly referred to as the belief or information state.

For robotics problems, we include the robot's own past actions as observed variables in the model. Figure 25.7 shows the notation used in this chapter: Xt is the state of the environment (including the robot) at time t, Zt is the observation received at time t, and At is the action taken after the observation is received.

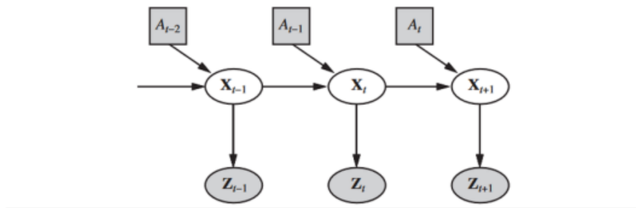


Figure 25.7: Robot perception can be viewed as temporal inference from sequences of actions and measurements.

We would like to compute the new belief state, P(Xt+1 | z1:t+1, a1:t), from the current belief state P(Xt | z1:t, a1:t−1) and the new observation zt+1.

Thus, we modify the recursive filtering equation to use integration rather than summation:

P(Xt+1 | z1:t+1, a1:t) = αP(zt+1 | Xt+1) ∫ P(Xt+1 | xt, at) P(xt | z1:t, a1:t−1) dxt .

This equation states that the posterior over the state variables X at time t + 1 is calculated recursively from the corresponding estimate one time step earlier. This calculation involves the previous action at and the current sensor measurement zt+1.The probability P(Xt+1 | xt, at) is called the transition model or motion model, and P(zt+1 | Xt+1) is the sensor model.

Localization and mapping:
Localization is the problem of finding out where things are—including the robot itself.

Knowledge about where things are is at the core of any successful physical interaction with the environment.
To keep things simple, let us consider a mobile robot that moves slowly in a flat 2D world. Let us also assume the robot is given an exact map of the environment. The pose of such a mobile robot is defined by its two Cartesian coordinates with values x and y and its heading with value θ, as illustrated in Figure 25.8(a). If we arrange those three values in a vector, then any particular state is given by Xt = (xt, yt, θt) .

RAMAIAH
Institute of Technology

For small time intervals **Δt**, with translational velocity **vt** and a rotational velocity **ωt**.
$\hat{\mathbf{X}}$ refers to a deterministic state prediction

$$\hat{\mathbf{X}}_{t+1} = f(\mathbf{X}_t, \underbrace{v_t, \omega_t}_{a_t}) = \mathbf{X}_t + \begin{pmatrix} v_t \Delta t \cos\theta_t \\ v_t \Delta t \sin\theta_t \\ \omega_t \Delta t \end{pmatrix} .$$

with mean f(Xt, vt, ωt) and covariance Σx

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_t, v_t, \omega_t) = N(\hat{\mathbf{X}}_{t+1}, \boldsymbol{\Sigma}_x) .$$

This probability distribution is the robot's motion model. It models the effects of the motion at on the location of the robot.
The exact prediction of the observed range and bearing would be

$$\hat{\mathbf{z}}_t = h(\mathbf{x}_t) = \begin{pmatrix} \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2} \\ \arctan \frac{y_i - y_t}{x_i - x_t} - \theta_t \end{pmatrix} .$$

RAMAIAH
Institute of Technology

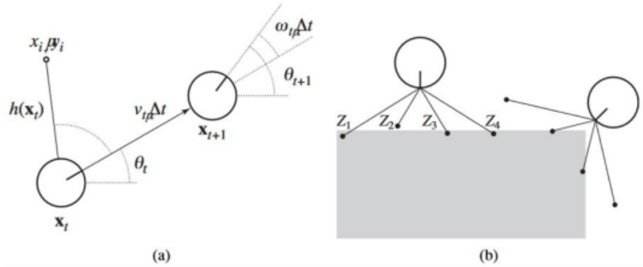


Figure 25.8 (a) A simplified kinematic model of a mobile robot. The robot is shown as a circle with an interior line marking the forward direction. The state $\mathbf{x}_t$ consists of the (x_t, y_t) position (shown implicitly) and the orientation θ_t. The new state $\mathbf{x}_{t+1}$ is obtained by an update in position of $v_t \Delta_t$ and in orientation of $\omega_t \Delta_t$. Also shown is a landmark at (x_i, y_i) observed at time t. (b) The range-scan sensor model. Two possible robot poses are shown for a given range scan (z_1, z_2, z_3, z_4). It is much more likely that the pose on the left generated the range scan than the pose on the right.

noise distorts our measurements. To keep things simple, one might assume Gaussian noise with covariance Σz, giving us the sensor model

$$P(\mathbf{z}_t \mid \mathbf{x}_t) = N(\hat{\mathbf{z}}_t, \mathbf{\Sigma}_z) .$$

A different sensor model is used for an array of range sensors, each of which has a fixed bearing relative to the robot. Such sensors produce a vector of range values zt = (z1,...,zM). Given a pose xt, let zˆj be the exact range along the jth beam direction from xt to the nearest obstacle. As before, this will be corrupted by Gaussian noise. Typically, we assume that the errors for the different beam directions are independent and identically distributed, so we have

$$P(\mathbf{z}_t \mid \mathbf{x}_t) = \alpha \prod_{j=1}^{M} e^{-(z_j - \hat{z}_j)/2\sigma^2} .$$

Figure (b) shows an example of a four-beam range scan and two possible robot poses, one of which is reasonably likely to have produced the observed scan and one of which is not. Comparing the range-scan model to the landmark model, we see that the range-scan model has the advantage that there is no need to identify a landmark before the range scan can be interpreted; indeed, in Figure (b), the robot faces a featureless wall. On the other hand, if there are visible, identifiable landmarks, they may provide instant localization.

Different methods for localization are:

- Monte Carlo Localization (or MCL)
- Extended Kalman Filter (or EKF)
- Simultaneous Localization And Mapping(SLAM).

RAMAIAH
Institute of Technology

MONTE CARLO LOCALISATION

Localization using particle filtering is called **Monte Carlo localization**, or MCL.
The gist of the particle filtering algorithm is that it uses set of particles (also called samples) to represent the posterior distribution of a stochastic process given some noisy or partial observations. For this to work, we need to supply an appropriate motion and sensor model. The robot requires a complete map in this case. The algorithm's working is demonstrated in figure 25.10, and the algorithm itself in 25.9. As is visible, the algorithm starts off unsure of its location in the map, and gradually as it moves, it becomes more and more informed of where it can be, and the number of particles(no of possible places it can be based on past observations) reduces.

This is visible in the progression of images. In the first image, the particles are uniformly distributed based on the prior, indicating global uncertainty about the robot's position. In the second image, the first set of measurements arrives and the particles form clusters in the areas of high posterior belief. In the third, enough measurements are
available to push all the particles to a single location. Eg: Neural Radiance Fields (NeRF)

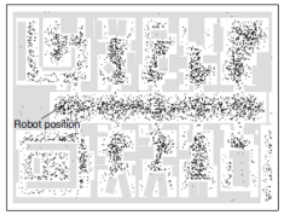
RAMAIAH
Institute of Technology

function MONTE-CARLO-LOCALIZATION(a, z, N, $P(X'|X, v, \omega)$, $P(z|z^*)$, m) **returns** a set of samples for the next time step
- **inputs**: a, robot velocities v and ω
  - z, range scan $z_1, \ldots, z_M$
  - $P(X'|X, v, \omega)$, motion model
  - $P(z|z^*)$, range sensor noise model
  - m, 2D map of the environment
- **persistent**: S, a vector of samples of size N
- **local variables**: W, a vector of weights of size N
  - S', a temporary vector of particles of size N
  - W', a vector of weights of size N

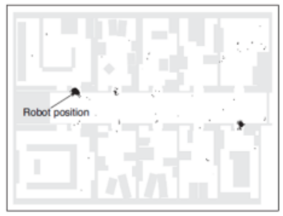
```
if S is empty then        /* initialization phase */
    for i = 1 to N do
        S[i] ← sample from P(X_0)
    for i = 1 to N do    /* update cycle */
        S'[i] ← sample from P(X'|X = S[i], v, ω)
        W'[i] ← 1
        for j = 1 to M do
            z* ← RAYCAST(j, X = S'[i], m)
            W'[i] ← W'[i] · P(z_j | z*)
    S ← WEIGHTED-SAMPLE-WITH-REPLACEMENT(N, S', W')
return S
```

Figure 25.9 A Monte Carlo localization algorithm using a range-scan sensor model with independent noise.

RAMAIAH
Institute of Technology

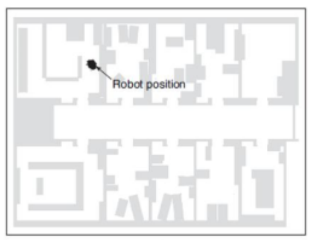


(a)



(b)

RAMAIAH
Institute of Technology



(c)

Figure 25.10 Monte Carlo localization, a particle filtering algorithm for mobile robot localization. (a) Initial, global uncertainty. (b) Approximately bimodal uncertainty after navigating in the (symmetric) corridor. (c) Unimodal uncertainty after entering a room and finding it to be distinctive.

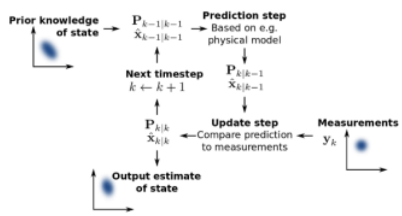
RAMAIAH
Institute of Technology

RAMAIAH
Institute of Technology

VIA KALMAN FILTER

Kalman filters can also be used to localize. The gist of Kalman filters is that it uses an algorithm that uses a series of measurements captured over time with noise, and outputs a probability distribution for the next measurement as well as other variables.

A Kalman filter represents the posterior $\mathbf{P}(\mathbf{X}t \mid \mathbf{z}1{:}t, a1{:}t-1)$ by a Gaussian. The mean of this Gaussian will be denoted $\boldsymbol{\mu}t$ and its covariance $\boldsymbol{\Sigma}t$. The main problem with Gaussian beliefs is that they are only closed under linear motion models f and linear measurement models h. For nonlinear f or h, the result of updating a filter is in general not Gaussian.



RAMAIAH
Institute of Technology

Thus, localization algorithms using the Kalman filter **linearize** the motion and sensor models. Linearization is a local approximation of a nonlinear function by a linear function. Figure 25.11 illustrates the concept of linearization for a (one-dimensional) robot motion model. On the left, it depicts a nonlinear motion model f(**x**t, at) (the control at is omitted in this graph since it plays no role in the linearization). On the right, this function is approximated by a linear function ˜ f(**x**t, at). This linear function is tangent to f at the point $\boldsymbol{\mu}t$, the mean of our state estimate at time t.

Such a linearization is known as taylor expansion. A Kalman filter that linearizes f and h via Taylor expansion is called an **extended Kalman filter** (or EKF). Figure 25.12 shows a sequence of estimates of a robot running an extended Kalman filter localization algorithm. As the robot moves, the uncertainty in its location estimate increases, as shown by the error ellipses. Its error decreases as it senses the range and bearing to a landmark with known location and increases again as the robot loses sight of the landmark.

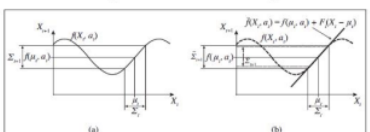


Figure 25.11 One-dimensional illustration of a linearized motion model: (a) The function f, and the projection of a mean μ_t and a covariance interval (based on Σ_t) into time $t+1$. (b) The linearized version is the tangent of f at μ_t. The projection of the mean μ_t is correct. However, the projected covariance $\tilde{\Sigma}_{t+1}$ differs from Σ_{t+1}.

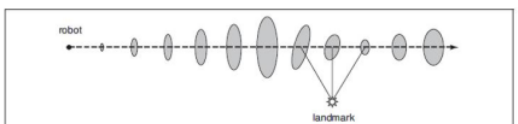
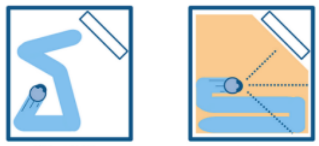


Figure 25.12 Example of localization using the extended Kalman filter. The robot moves on a straight line. As it progresses, its uncertainty increases gradually, as illustrated by the error ellipses. When it observes a landmark with known position, the uncertainty is reduced.

RAMAIAH
Institute of Technology

S.L.A.M

In some situations, no map of the environment is available. Then the robot will have to acquire a map. This is a bit of a chicken-and-egg problem: the navigating robot will have to determine its location relative to a map it doesn't quite know, at the same time building this map while it doesn't quite know its actual location. This problem is important for many robot applications, and it has been studied extensively under the name **simultaneous localization and mapping**, abbreviated as **SLAM**.
SLAM problems are solved using techniques based on probability, since nothing is sure other than absolute motion of the robot w.r.t the ground.
SLAM problems are solved using many different probabilistic techniques, including the extended Kalman filter discussed above. Using the EKF is straightforward: just add the locations of landmarks to the state vector. EKF updates scale quadratically, thus it'll be a problem for multiple landmarks in the same models. So in most situations small models are made and their relative positioning is obtained by the same algorithm, which'll make it easier for computation. Eg: Photogrammetry



Without SLAM:
Cleaning a room randomly.

With SLAM:
Cleaning while understanding the room's layout.

RAMAIAH
Institute of Technology

OTHER SENSES OF PERCEPTION

- Not all of robot perception is about localization or mapping. Robots also perceive the temperature, odors, acoustic signals, and so on. Many of these quantities can be estimated using variants of dynamic Bayes networks.
- All that is required for such estimators are conditional probability distributions that characterize the evolution of state variables over time, and sensor models that describe the relation of measurements to state variables.
- The trend in robotics is clearly towards representations with well-defined semantics. Probabilistic techniques outperform other approaches in many hard perceptual problems such as localization and mapping.
- However, statistical techniques are sometimes too cumbersome, and simpler solutions may be just as effective in practice.

MACHINE LEARNING IN ROBOTIC PERCEPTION

- Machine learning plays an important role in robot perception. This is particularly the case when the best internal representation is not known.
- One common approach is to map high- dimensional sensor streams into lower-dimensional spaces using unsupervised machine learning methods. Such an approach is called low-dimensional embedding.
- Another machine learning technique enables robots to continuously adapt to broad changes in sensor measurements.
- Our perception quickly adapts to the new lighting conditions, and our brain ignores the differences.
- Adaptive perception techniques enable robots to adjust to such changes.
- Methods that make robots collect their own training data are called self- supervised.

(a) (b) (c)

Figure 25.13 Sequence of "drivable surface" classifier results using adaptive vision. In (a) only the road is classified as drivable (striped area). The V-shaped dark line shows where the vehicle is heading. In (b) the vehicle is commanded to drive off the road, onto a grassy surface, and the classifier is beginning to classify some of the grass as drivable. In (c) the vehicle has updated its model of drivable surface to correspond to grass as well as road.

THANK YOU

BIOMETRIC_AUTHENTICATION_FOR_ATMs.docx

ORIGINALITY REPORT

11%	10%	5%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	medium.com Internet Source	5%
2	vis-www.cs.umass.edu Internet Source	2%
3	Submitted to University of Sheffield Student Paper	1%
4	www.antiessays.com Internet Source	1%
5	isindexing.com Internet Source	<1%
6	sersc.org Internet Source	<1%
7	stdhomes.ieu.edu.tr Internet Source	<1%
8	link.springer.com Internet Source	<1%
9	Submitted to University of West London Student Paper	<1%

10	W. Zhao. "Face recognition", ACM Computing Surveys, 12/1/2003 Publication	<1 %
11	Pranjal Saikia, Madhabi Konwar, Shyamalee Patar, Baishali Mahanta, Pranjit Barman, Lakhya Jyoti Borthakur. "Electrochemical performance of rGO/AFCNT composite and fabrication of supercapacitor electrode", Materials Today Communications, 2022 Publication	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		