DRISHTI: REAL TIME APPLICATION FOR VISUALLY IMPAIRED PEOPLE

Mr. Dhruv Pithwa[1], Mr. Bhavin Mehta[2], Mr. Priyam Vora[3], Prof. Arjun Jaiswal[4]

[1,2,3] *Student, Information Technology, Dwarkadas J. Sanghvi College of Engineering (India)*
[4] *Asst. Prof. Information Technology, Dwarkadas J. Sanghvi College of Engineering (India)*

Abstract - *In India, about 62 million people are blind or visually impaired (as of April 2019). Nearly 54 million of these people have poor vision, and 8 million are blind. Visual impairment also contributes to a sense of reliance on sighted people. As a result, there is a feeling of restlessness, and such people often separate themselves. The inability to move easily is a major problem of visual impairment. There are also other issues that these people face on a daily basis. Both types of people can encounter navigation problems at some stage during their daily lives. When it comes to visually disabled individuals, they face a lot of difficulties while travelling. Several systems have been developed in the past with the aim of assisting visually disabled people and improving their quality of life. Unfortunately, the majority of them were either prohibitively costly or poorly designed. Educated dogs and white canes are the easiest and most economical navigation tools available to them. Despite their popularity, these tools cannot provide the blind with all of the knowledge and functionality necessary for secure mobility that are accessible to normal people. Our system's main goal is to make it easier for blind and visually disabled people to travel freely. In addition to these issues, the proposed framework seeks to address two other issues that BVI (Blind Visually Impaired) people face, namely facial recognition and currency denomination detection.*

Key Words: Mobile Application, 3D audio enabled Navigation System, Face Recognition System, Currency Detection System

1. INTRODUCTION

BVI people face numerous hurdles in their day-to-day functioning. In this proposed paper we aim to address the following enumerated hurdles of - 1) outdoor navigation 2) face detection and recognition 3) currency denomination detection. The proposed solution provides modifications on some of the significant research work done on Navigation, face recognition and currency recognition for visually impaired all over India. 3D Voice navigation and GPS tracking in real time is of done for visually handicapped through this system.

1.1 Outdoor Navigation

BVI face many challenges in their daily lives. We plan to resolve the following enumerated challenges in this proposed paper: 1) navigation in open spaces 2) Facial recognition and detection 3) Identification of currency denominations. The proposed solution incorporates changes to some of the most important studies on navigation, face recognition, and currency recognition for visually disabled people across India. This system provides 3D voice navigation and real-time GPS monitoring for visually impaired people. Communication technologies have advanced at a breakneck pace over the last decade. Wireless networks with high bandwidth and low latency are commonly used around the world, providing added advantages in people's daily lives and assisting in making them more mobile and versatile. Mobile devices are also becoming easier to obtain and afford as a result of a large and rising demand for portable appliances [3]. Collaboration between existing communication technologies and accessible smart-phone devices results in more dependable and efficient products across all industries, including healthcare [4]. Despite the above, it is safe to say that visually disabled people are still restricted in their mobility in the vast majority of cases. As a result, the development of special facilities will not only improve the mobility of visually disabled people, but also give them a greater sense of security in their daily lives. To navigate, visually impaired or blind people rely on their prior knowledge of an area, with the assistance of guide dogs or white canes. [2] As a result, they struggle to achieve the desired degree of mobility and context awareness, especially in unfamiliar environments.

1.2 Facial Recognition

To recognise faces, FaceNet, a Google algorithm, is used. FaceNet is a centralised embedding that can be used for tasks such as facial recognition, authentication, and clustering. It transforms each face image into a Euclidean space with distances contributing to face resemblance, i.e., a person's image. [8] When tested image with images of every other person in the database, B will be put closer to all other images of person B.

FaceNet differs from other techniques in that instead of relying on a bottleneck layer for identification or verification, it learns the mapping from the images and produces embeddings. Following the creation of the embeddings, the remaining tasks, such as verification and identification, can be accomplished using standard techniques for that domain, with the newly formed embeddings serving as the function vector.

1.3 Currency Recognition

Deep Learning Convolution dependent Xception architecture with Depthwise separable convolution is used to detect currency. In terms of computational power, Xception Net is an improvement on InceptionNet. Extreme

Inception is what Xception stands for. Rather than an InceptionNet, the Xception architecture shown in the image above is more akin to a ResNet. Inception Net v3 is outperformed by Xception Net.

The distinction between Inception Net and Xception Net is that Inception Net performs regular convolutional operations, while Xception Net performs Depthwise Separable Convolutional operations [11]. Depthwise Separable Convolutions differ from normal convolutions in that, in a normal Conv2D layer, we can use any number of filters in the Conv layer for an input of (32, 32, 3) picture. Each of those filters will be applied to all three channels, with the number of all corresponding values as the output. In Depthwise separable convolutions, however, each channel has only one convolution kernel. As a result, we may reduce the computational complexity by conducting Depthwise Separable Convolutions since each kernel is only two dimensional and only convolutes over one channel. We can do this in Keras by using the DepthwiseConv2D layer.[6]

In this paper, we propose the Drishti navigation method, which is transparent, available, and extensible, paving the way for a new paradigm for visually impaired people in navigation, individual detection, and currency denomination detection. The aim of our study is to create a mobile application that will help vision-impaired people navigate safely and independently while also allowing them to quickly identify people around them. The solution relies on the use of geolocation-based navigation technologies as well as computer vision technology for detection and recognition.

The user will be given an Android-based application that recognises voice commands and allows them to navigate around different locations as well as recognise friends and family members. The proposed application is realistic and provides users in outdoor settings with near-accurate position guidance. In addition, the machine learning model created for detection purposes has a high level of accuracy in terms of identification. Over all, the device aims to raise the bar for the standard of navigation and knowledge of the environment and people around the visually impaired.

2. LITERATURE SURVEY

2.1 Outdoor Navigation

2.1.1 Navigation System using RFID & MICHIBIKI Satellite

In order to build a navigation system for visually disabled people, they created a high-accuracy positioning system using both UHF band RFID and MICHIBIKI in this research. The measurement tests for various RFID tags were carried out with the assumption that they would be installed in real-world environments. They also propose a position estimation method based on RFID and MICHIBIKI position data. As a result, they have verified that the positioning system can be used to navigate visually disabled people both indoors and outdoors with adequate precision.

2.1.2 ARIANNA

This system is suitable for both indoor and outdoor use and is built on a combination of dead-reckoning and computer vision algorithms. They showed how special reference signals, such as color tapes intended to replace tactile paving, can be used to transform the smartphone camera into an additional sensor. Indeed, it is possible to determine the user's heading and velocity by processing the camera's images, while detecting special landmarks, such as corners or visual markers, will completely reset tracking errors in close proximity to known coordinates. In measuring heading, computer vision functions are more precise than gyroscopes, with a median error of about 5 degrees [4]. Computer vision functions that measure velocity, on the other hand, only operate well when the user's movements do not include significant rotations. The smartphone IMU, on the other hand, can detect rotations and provides an alternative velocity estimate, so the opportunistic combination of all the measurements greatly increases the tracking system accuracy.

2.1.3 Research Experiment

The use of GPS mobile applications for blind pedestrian navigation was investigated in this paper [5]. The focus of future research will be on increasing user trust in GPS-based applications in terms of accuracy and environmental perception. The study aided in the analysis and comprehension of blind people's use of technology for navigation. The study also highlighted the lack of adequate knowledge provided to the blind while navigating in an unfamiliar setting. It was also discovered that using the correct phrases and units is extremely important, as all of the participants heavily relied on the distance measuring units of the apps while being led by GPS-based apps. To feel secure and protected when navigating, all of the blind participants requested precise locations without any errors. As a result, the effectiveness of current applications can be improved. The lack of comprehension of the GPS's synthesized voice triggered the navigational errors in the distance. Due to inadequate information provided by the applications, the pace of blind people when navigating using GPS slowed as well.

2.2 Face Recognition

2.2.1 Classical Face Recognition Algorithms

Advantages: -

- Emphasizes the manifold's local structure.
- Face is projected onto a linear subspace spanned by the eigenface images using these methods.
- Due to the fact that the distance from face space is orthogonal to the plane of the mean image, it can be easily converted to Mahala Nobis distances using probabilistic interpretation.

Disadvantages: -

- When broad variations in lighting, facial expressions, and other factors occur, these methods can fail to accurately reflect faces.
- In this case, using kernel-based nonlinear methods does not result in a substantial improvement over linear method.
- LLE, LLP, and LBP were created to provide a quick and efficient way to explain neighboring changes in face description. In DCV and SVM-based methods, subspace approaches were used. NPP and ONPP methods are concerned with preserving local structure between samples.
- The issue is that it's still unclear how to choose the size of the neighborhood or assign optimal values to it.

2.2.2 Artificial Neural Network

Advantages: -

- The radial basis function (RBF) Nonnegative matrix factorization is naturally combined with artificial neural networks.
- There are also other approaches for process simplification involving the native linearization function of ANNs and increased computation speed.
- Ideal solution, particularly for recognizing partial distortion and occlusion in face images.

Disadvantages: -

- The biggest drawback of this strategy is that it necessitates a larger number of training samples (instead one or limited number).
- It's unreliable in the same way that other statistically dependent approaches are.

2.2.3 Gabor Wavelets

Advantages: -

- Gabor wavelets have the advantage of being able to capture important visual properties such as spatial localization, orientation selectivity, and spatial frequency.
- This method is preferred in a variety of biometrics applications.

Disadvantages: -

- Since the face image is convolved with a bank of Gabor filters, the Gabor feature space has a substantially high dimensionality, which is a disadvantage of Gabor-based approaches.
- The method is computationally expensive and thus unsuitable for real-time applications.
- Simplified Gabor characteristics are also susceptible to lightning variations.

2.2.4 Face Descriptor Methods

Advantages: -

- The main goal of designing image descriptors is to discover the most discriminant local features that reduce the difference between images of the same person while increasing the difference between images of different people.
- These approaches are discriminative and resistant to changes in lighting and voice. They have descriptors that are lightweight, simple to extract, and highly discriminative.

Disadvantages: -

- During the descriptor extraction point, the approach is computationally intensive, but it encourages simplicity and efficiency in comparison to online applications.

2.2.5 Video Based Recognition

Advantages: -

- The approach's key benefit is the ability to use the redundancy found in video to enhance still image systems.

Disadvantages: -

- Relatively little research has been done on this. Problems with calculating the resemblance of two (or more) images are multiplied.

3. PROPOSED SYSTEM

Drishti is an application designed specifically while keeping in mind a visually impaired audience. The scope of the project includes helping the visually impaired navigate freely without any dependence in outdoor spaces. The major aspect of this project is to guide them in outdoor spaces through use of 3d audio and some sort of feedback. Other major feature includes helping them recognize people around them through facial recognition and save detail of the known one for better audio description of the person the next time. Next feature includes currency denomination detection. Right now, the focus will be on Indian denomination albeit it will be expandable.

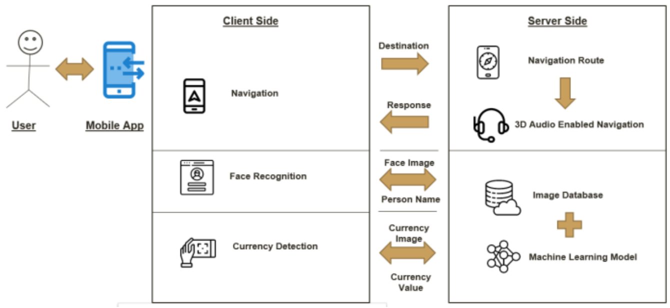


Figure 1 - System Architecture

User will be provided with a mobile application through which all the features will be accessible. In the mobile application there are going to be four main features, which are Outdoor Navigation, Face Recognition and Currency Denomination Detection. For navigation, the user will have to mark its destination location and the shortest route will be provided to the user. Along with this the app will continuously interact with server for providing 3D audio enabled navigation. For face recognition and currency detection, user will have to click a picture and submit it, and the result will be provided to the user such as the name of the person or the value of the currency.

4. SYSTEM IMPLEMENTATION

4.1 Outdoor Navigation

Geolocation Technology will be used to mark the destination for the user and navigate the user through the shortest available path. A 3D audio assistance will provide continuous response to the user about the direction on which user has to walk. The 3D audio will be rendered dynamically to keep user updated about the direction. An assistance/notification about nearby places will inform users about the nearby landmarks and surroundings.

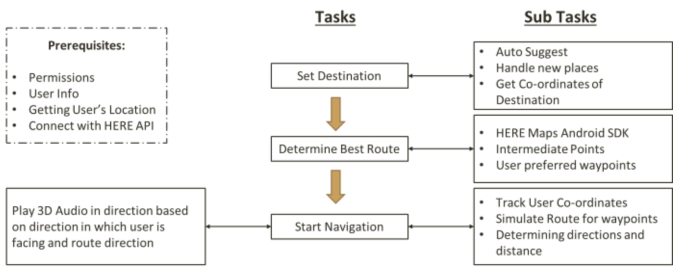


Figure 2 - Outdoor Navigation Tasks

In order to initiate outdoor navigation, there are some prerequisites that needed to be ensured, which includes getting permissions from user such as navigation, mic and storage, taking the required user information within the application and identifying the user's current location and establishing connection with HERE maps API. Once the prerequisites are completed, the user will be asked to set up the destination which involves tasks such as auto suggesting the user based on the keywords entered by the user and getting the location coordinated of the destination once the destination is selected. The next task is identifying the shortest route available for navigation, which includes communication with the HERE maps API and handling user preferred intermediate points, such as any preferred places through which the user wants to navigate. Once the route is identified, the user will be informed about the direction in which user has to move forward with the help of 3D audio and will be kept informed about the surroundings while navigating towards the destination. The user's location coordinates will be tracked throughout the route along with the user heading bearing in order to calculate 3D audio angle based on user's face direction. This way with the help of 3D audio and Geolocation technology, outdoor navigation will work.

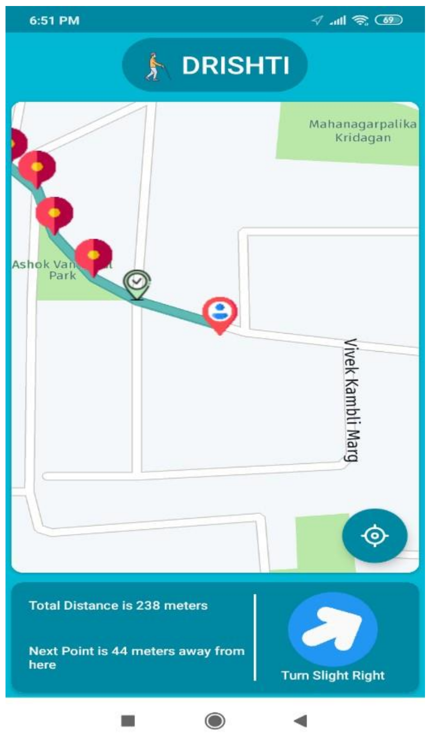


Figure 3 - Outdoor Navigation using 3D Audio

4. 2 Facial Recognition

To recognise faces, FaceNet, a Google algorithm, is used. FaceNet is a centralised embedding that can be used for tasks such as facial recognition, authentication, and clustering. It transforms each face image into a Euclidean space with distances contributing to face resemblance, i.e., a person's image. [8] When tested image with images of every other person in the database, B will be put closer to all other images of person B. FaceNet differs from other techniques in that instead of relying on a bottleneck layer for identification or verification, it learns the mapping from the images and produces embeddings. Following the creation of the embeddings, the remaining tasks, such as verification and identification, can be accomplished using standard techniques for that domain, with the newly formed embeddings serving as the function vector. When a user submits an image to a face recognition system, the system should check to see if the input person's image already exists in the user's database; if it does, the user should be informed and given the name of the identified person; otherwise, the user should be asked to add the unknown person to his or her database.

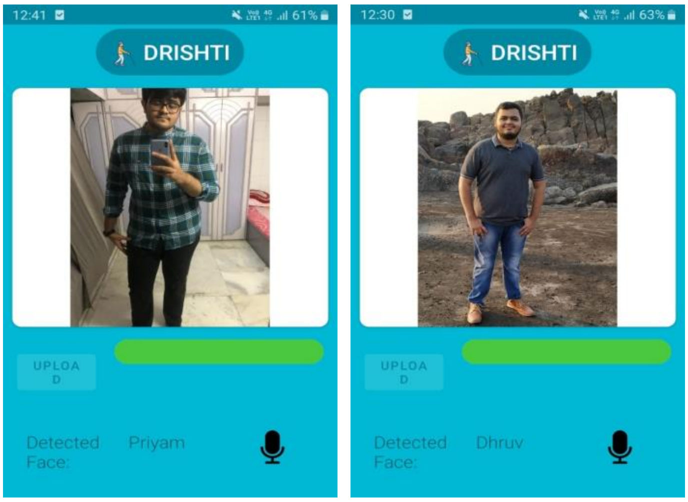


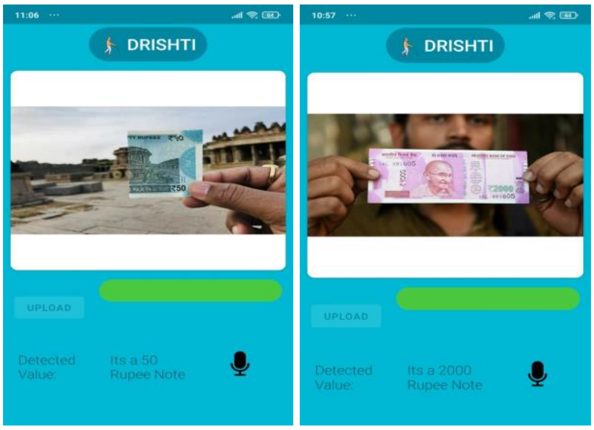
Figure 4 - Face Recognition

Upon giving input image to face recognition system, the system should check whether the input person's image is there in user's database, if it is there then user should be notified with the known person's name else the user should be prompted to add that unknown person into his/her database.

4.3 Currency Recognition

Deep Learning Convolution dependent Xception architecture with Depthwise separable convolution is used to detect currency. In terms of computational power, Xception Net is an improvement on InceptionNet. Extreme Inception is what Xception stands for. [nine] Rather than an InceptionNet, the Xception architecture shown in the image above is more akin to a ResNet. Inception Net v3 is outperformed by Xception Net. The distinction between Inception Net and Xception Net is that Inception Net performs regular convolutional operations, while Xception Net performs Depthwise Separable Convolutional operations.

Depthwise Separable Convolutions differ from normal convolutions in that, in a normal Conv2D layer, we can use any number of filters in the Conv layer for an input of (32, 32, 3) picture. Each of those filters will be applied to all three channels, with the number of all corresponding values as the output. In Depthwise separable convolutions, however, each channel has only one convolution kernel. As a result, we can reduce the computational complexity by conducting Depthwise Separable Convolutions since each kernel is only two dimensional and only convolutes over one channel [6]. We can do this in Keras by using the DepthwiseConv2D layer.



When a currency recognition system receives an input image, it should process it, compare it to standard currency notes, and make a result based on the comparison.

5. BENEFITS TO THE SOCIETY

- With the help of Project Drishti, Visually Impaired people will be able to carry out their routine activities easily: Navigation, Currency Detection & Facial Recognition.
- Visually Impaired people will be able to navigate easily and with independence in outdoor environments and they will also be able to identify various routes to reach their desired destinations.
- 360-degree orientation will enable users to point their phone in their direction of choice to discover what lies ahead.
- Visually Impaired people will be able to identify their friends and family and also register new people in the list of friends.
- Difficulty in recognizing currency while carrying out transactions/ bill payments will be mitigated through our system.

6. FUTURE SCOPE

- To navigate indoor spaces with ease using Bluetooth based transmitters called beacons working on principle of Bluetooth Low Energy (BLE). [10]
- In order to identify the objects in the surrounding area, object and obstacle detection module can be included within Drishti.
- Visually Impaired People find it difficult in reading text through any material thus feature like image to text to speech can be included.
- Scene descriptor - Describe the objects and scenes nearby for Visually Impaired people.
- Realtime GPS tracking of Visually Impaired People to keep the near ones apprised of their location.

Figure 5 - Currency Recognition

- Barcode descriptor - Describe product by scanning the barcode/QR code on the product.
- Expand currency detection module by integrating international currencies such as the dollar, pound et al.

7. CONCLUSION

While designing the system `Drishti` we have taken into count many parameters such as time delay, cost feasibility, good accuracy, less computation, flexibility, etc. Thus, considering all these parameters we have come to some methodology/approach we are going to use to meet all these metrices. For Outdoor Navigation, we are going to use 3d Audio based Geo-Technology service which have advantages like better navigation of directions, no external hardware requirement, etc. For Face Identification, we are going to follow face descriptor-based methodology which is practical when it comes to real world applications as it not only requires less computation but also it has very good accuracy as far as our system Drishti is concerned. For Currency Denomination, it will be feature extraction-based approach where pre-trained model will be used and tested on the user's input.

REFERENCES

[1] Farooq Shaikh, Mohammad Abbas Meghani, Vishal Kuvar, Shiburaj Pappu “Wearable Navigation and Assistive System for Visually Impaired”

[2] Rabia Afzal Minhas, Ali Javed “X-EYE: A Bio-smart Secure Navigation Framework for Visually Impaired People”

[3] Arnav Bakshi, Arjun Bakshi “Energy and Cost Efficient Navigation Technique for the Visually Impaired”

[4] Priscilla Ebenezer R, Vishnu Priya M, B. Nivetha “GPS Navigation with Voice Assistance and Live Tracking for Visually Impaired Travelers”

[5] Israh Akbar, Ahmed Fatzilah Misman “Research on Semantics Used in GPS Based Mobile Phone Applications for Blind Pedestrian Navigation in an Outdoor Environment”

[6] M. Laavanya, V. Vijayaraghavan "Real Time Fake Currency Note Detection using Deep Learning"

[7] Vidhi Roy, Gangey Mishra, Rahul Mannadiar, Sushant Patil "Fake Currency Detection Using Image Processing" retrieved from

[8] Sujata G. Bhele, V. H. Mankar "Review on Face Recognition Techniques"

[9] Andrew Mackey, Petros Spacho, Konstantinos N. Plataniotis “Enhanced Indoor Navigation System with Beacons and Kalman Filters” retrieved from

[10] Seyed Ali Cheraghi, Vinod Namboodiri “Guide Beacon: Beacon-Based Indoor Wayfinding for the Blind, Visually Impaired, and Disoriented” retrieved from

[11] YOLO-v3 Based Currency Detection and Recognition System for Visually Impaired Persons