

Fabricating Modish Mobile Application for a Visionless Citizen to Act as a Visioned Citizen

Jananee V*, Shalini P, Dhivya S

Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.

Article Info

Article history:

Received Nov 7, 2017

Revised Jan 27, 2018

Accepted Feb 26, 2018

Keywords:

Gesture to speech Algorithm

Hidden Markov Model

Text to speech conversion

Viterbi Algorithm

ABSTRACT

Technology has its wide range of applications in every field potentially even on the Mobile technology. One of the technologies which accomplish the blind person in android mobile is Virtual Reality. In spite of virtual reality is engaged to effectuate the operations, the blind person attention is one of the most important statistics. Although mobile devices include operative features available for amaurotic users, the user interface of the greater part of the mobile apps is designed for sighted people. If they carry out any mistakes to use the apps it may lead to a wrong call. So, one may think of a technology that diminishes the anxiety of a blind person for using apps. The proposed system converts the text into audio for giving the directions to the blind person about the gestures inferred. For such conversion a technique called speech synthesizer is used. Assorted innovative tools are used in mobile phones. Blind people need to confide in normal person for creating and updating a contact. Our project incorporates the gestures from the blind people and confirms the gestures through voice. Gesture conversion is accomplished by using haptic technology. An amaurotic person can create a new contact and they can invoke the contact by using this voice confirmation. During calling if there is multiple contact list, caller setting are initialized and ask the preference from the blind people, progress the voice call.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Jananee V,

Department of Computer Science and Engineering,

Rajalakshmi Engineering College,

Chennai, Tamil Nadu, India - 602105.

Email: jananee.v@rajalakshmi.edu.in

1. INTRODUCTION

Cell phones are a vital part of today's life. Many of us wish to make a call or send a message at any moment from any place. For visually groped users voice based contact lists are provided with many cell phones. They can select contacts through voice and make calls when required. Speech Recognition and Conversion will be the intrinsic part of the application. Android actually provides support for those groups which are really not noticed by many. Coming to differently abled people, they face more problems than the usual being. It will always be an ecstasy for them to enjoy as normal being with all such factors. The application is focused on differently abled set of humans, who may not be in the situation of using mobile phones.

In less than a couple of years, gesture based interaction has become a standard on the best part of the mobile devices. It is an upcoming area of research since touch displays are more and more present in our daily life. Touch screens provide a great flexibility and a direct access, but on the other hand, less accessible to visually impaired and blind users. The goal of our research work is to promote low vision users the interaction with device that uses this kind of displays. So, this paper deals about design. Design of touch based mobile apps fit for visually impaired people [1]. Haptic Technology takes superiority of user's

sensitivity of touch by applying extortion, oscillation and direction. Haptic refers to sensing and direction through touch.

As touch screens have become primary, it is pivotal that touch screen based interfaces be usable by people with all abilities, including blind and visually impaired people [2]. Unlike fully accredited or sighted people, blind people cannot view messages displayed on the smart phones or cannot use the basic operations like calling, messaging, etc. Hence, communication via mobile device is a challenge for blind users, who often confront severe accessibility issues. The main problems are due to the lack of hardware keys, making it tough to quickly reach an area or activate functions. A touch screen has no unique reference points distinguishable by feel. So, a blind user feels hard to figure out where he is positioned exactly or to find a specific item/function. Hence, we disagree that operative touch screen interfaces can be enhanced greatly, if the designers can realize how blind people actually use touch screens.

Haptic device acts as an input and output device, capturing user real administrations as an input and furnishing realistic touch sensations [3] as an output accommodated with onscreen actions. As technology advances and computer power emerges, haptic devices and properties expands and get more realistic. This technology has verified that implicit objects can also be touched, felt and inhibited. This technology must be made feasible for the fair cost and the haptic devices must be made smooth and easier to use.

Haptic technology [4] is extensively used in gaming, surgical simulation, medical training, military training in virtual environment, Robotics, Virtual arts and design, mobile devices, research and entertainment. Haptic application depends upon highly functional hardware and requires huge transformation power. Finally, it is ensured that the haptic technology is the result for communicating with the virtual environment.

Generate a message application for text to voice modification and conversely voice is enabled using on-demand language model interposition [5]. This application receives your message and acknowledges with voice notification by pronouncing the same. As a part of sending message, this application is liable for voice to text transference which is uttered by user, and again text to voice to review message. Text to Speech is also identifying new operations outward to the infirmity market. For instance, speech integration, combined with speech realization, confess for communication with mobile devices via common language processing interfaces.

Using NEW VISION, calls and messages can be made using pattern detection and the spot of the user can be fetched using Global Positioning System technology [6]. Furthermore, we start a text-to-speech interface and achieve through vibrations to comfort the usage of smart phones for the blind users. Also, other functionalities like calling, messaging, time, battery level etc. are made simple for the visually challenged users. Application like "Voice for Android", is implied for visually challenged. It is a global translator for mapping images to sounds. Other applications such as "Mobile Accessibility" have calling and messaging features, but they take voice as input and are not very potent for Indian English accent.

The Text To Speech (TTS) [7] conversion with language translation is achieved for the mobile on android environment. It is a Natural Language Processing (NLP) module that affords easy communication for the person who cannot speak but can interact verbally. Also for the person who cannot perceive other regional languages can choose the language manually by using this application.

TTS conversion with language translator converts the normal language text into artificial formulation of human speech. This work changes the written text form to a phonemic representation. Later, converts the phonemic representation to waveforms that can be output as intonation sound. NLP is a field of human-computer synergy that makes a computer to understand and manipulate human language text or speech. Initially, get the input text in the English language. After getting the text, separation of the English words from the text is performed. Then, we execute the library lookup to get the phonetic equivalent of the text and arrange these entire phonetic equivalents in a series relevant to the text. Consequently, speech synthesis is achieved and the speech quality is retained.

The intention in transforming multiple algorithms such as 13 point feature extraction and 23 point feature extraction is to help revamp performance. For pattern processing two major accession such as online and offline processing were considered, out of which online refining was used as it is faster than offline processing and there is no need to redeem the pattern as image. This application through pattern pairing, gesture detection and voice messaging would make dialing and messaging from smart phones accessible and uncomplicated for visually impaired [8].

Khan et al. [9] Student-GLASS wearable is designed for smart camera device built with a powerful microcontroller that has the ability to see what we, normal people, are seeing, understands the user voice requests and supplies the relevant information using auditory feedback through an earphone. The device aims to improve the quality of life for the blind and visually impaired people and makes them understand their surroundings in a clear way as close as to a normal person at affordable costs.

Prakash et al. [10] implemented an approximation automated structure, called Filtered Wall (FW) and it filtered disposed of substance from OSN client substances. The goal is to utilize efficient classification procedure to stay away from overpowered by unsuccessful messages. In OSNs, content filtering can also be abused for a unique, more reactive.

In [11] explained integration of Adaptive Weight Ranking Policy (AWRP) with intelligent classifiers (NB-AWRP-DA and J48-AWRP-DA) via dynamic aging factor to improve classifiers power of prediction. The methods are used to choose the best subset of features. In [12] introduced a new framework called Fuzzy based contextual recommendation system for classification of customer reviews. It extracts the information from the reviews based on the context given by users. In [13] studied to identify the best classifiers for class imbalanced health datasets through a cost-based comparison of classifier performance. The unequal misclassification costs were represented in a cost matrix, and cost-benefit.

The first algorithm, the Baum – Welch scheme, is used to re – evaluate the model parameters. The second algorithm, the Viterbi method, is used to estimate the ability of an HMM at describing a particular observation independent of the domain for which HMMs are being applied. Visualization would stimulate a much better understanding of the system passage. For example, if an adventurer notices that the HMM parameters have already encountered, he can stop the training process. This type of visualization can be accomplished by exposing the parameter matrices as images, where the brightness of an image location resembles to the relative weight of an entry in the matrix. As another example, when a human sees not only the tough HMM for describing an example but also its strength relative to other HMMs, the label assigned to the example in recognition can be certified.

2. RESEARCH METHOD

Acceptance of contents through gesture plays a vital role. Still visually challenged people didn't reach the friendly procure at smart mobile stuff. This is a major contention where gestures are badly recognized by smart phones. There leads a connection gap between smart android app and the user. Gesture alone will not guide for a good access of smart phones. We need on app in such a way that there should be a recognition saying that the information / data requested by the user is acquired correctly. This type of communication gap can be blown-away by our paper. Visually impaired can access our android kit as a normal human being without one's help. Figure 1 explains about the design of proposed system in detail.

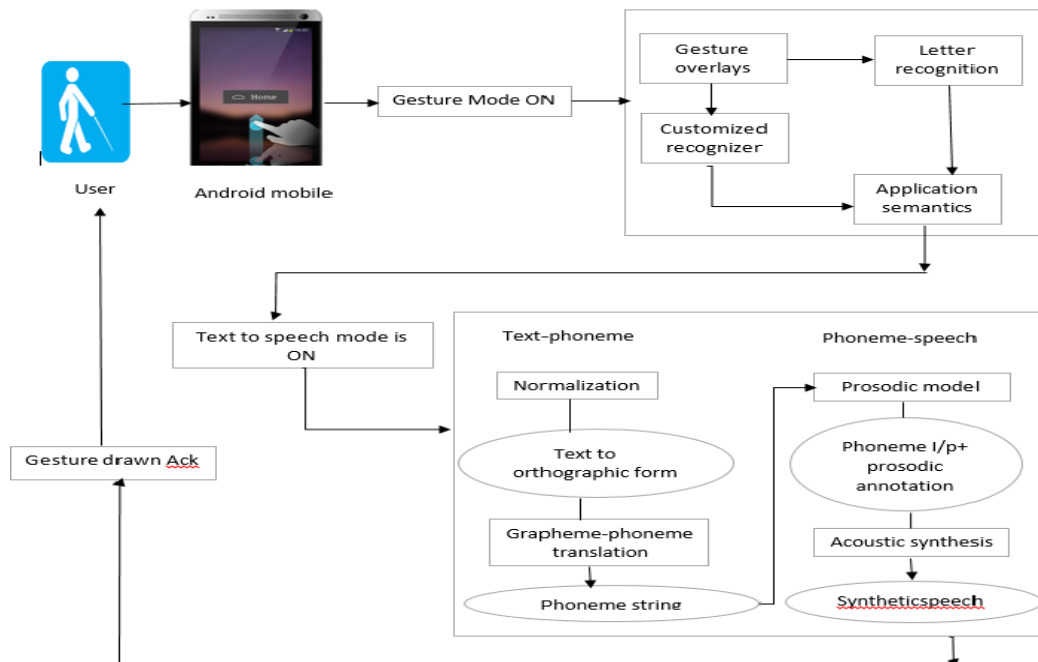


Figure 1. Design of Proposed System

Cell phones are small comfortable for us to carry them all around. With a cell phone, we can make phone calls, check our emails, surf the web or send text messages. Regrettably, we may have a hard time seeing our cell phone's screen in direct sunlight. Either we may get a glare, or we may not see anything at all. We can prevent these problems from occurring the next time you take your cell phone. Our Paper is focused on two types of mode. Gesture mode and Voice mode. This Paper speaks out the convenient and efficient usage of Smart Mobile Application for visually impaired people.

2.1. Multidimensional HMM

Gestures are converted into Sequential Symbols. HMM is a finite state, connected with Multiple Transitions. Each state has 2 probability sets. One is discrete output density function and the other is continuous output density function. Multidimensional gesture is one of the multi-path recognition process. Based on the time series gestures are classified into G (x,y,t). G: Gesture pattern drawn in and around X and Y axis.

S - Set of states

A: Transition Probability Matrix

B: Output Probability of discrete HMM

|A_{ij}|: Transition State from i and j

|B_j(x)|: x represents Continuous Observation

|B_j(OK)|

Ok: Discrete Observation Symbol

K: Random vector

“Discrete HMM”

A_{ij} ≥ 0, B_j(OK) ≥ 0, ∀ i, j, k

∑_j A_{ij} = 1, ∀ i, j

∑_k B(Ok) = 1, ∀ k

Λ = (A, B, π)

Π: Initial State of distribution

2.2. Multidimensional Approach

1. Gesture Defining

2. Drawing a pattern that is evolved in android mobile based on x, y axis at time t. Note down the pixels of the drawn pattern. Array Size is decided dynamically by retrieving screen resolution. Based on this, binarization is done specifying x(t): no of rows, y(t): no of columns

HMM Procedure

A multidimensional HMM has N distinct hidden states and M observable symbols

A: Transition state

B: Discrete output distribution

3. Collect Training Set

Raw data is pre-processed before training HMM. Training data is collected using STFT (Short Term Fourier Transform)

STFT Process: Segment the signal into narrow time intervals and take Fourier Transform for each signal. Each Fourier Transform is based on the Time Slice of the signal, providing time and frequency information.

$$\text{STFT}_f^u(t', u) = \int_t f(t)w(t - t') \cdot e^{-j2\pi ut} dt$$

t': Time Parameter

u: Frequency

f(t): Analyze signal

w(t-t'): Window Functioning centered at t to t'

STFT has time localization but no frequency localization

$$\text{STFT } f(t') = f(t') \cdot e^{-j2\pi ut'}$$

Gesture Recognition:

$$G^* = \arg \max p(\Lambda/O)$$

STFT Strategies:

- Choose a window function of finite length
- Place a window on top of the signal at t=0
- Truncate the signal

- Compute Fourier Transform for the truncated signal

1. Train HMM with Training data
Likelihood= $P(O/\Lambda)$ HMM based approach are stuffed with trained data
2. Evaluate gestures with trained data
Matching with trained sets, gestures are identified using Viterbi algorithm

2.3. Viterbi algorithm:

Viterbi is data and memory intensive procedure for matching the Minimum path clusters.

1. Decode a data sequence that has been encoded by finite state process
2. Viterbi Algorithm is optimal in maximum likelihood sense
3. Viterbi calculates a semi brute force estimate of likelihood for each path through trellis
4. Trellis: based on starting state for all possible sequence are gathered

- Calculating Trellis
- Finding Shortest path
- trace back
- Reorder output bits

LBG Algorithm:

LBG Algorithm splits the training vectors into $2, 4, \dots, 2^m$ partition and determines the centroid for each partition. It is refined iteratively by k-Means Clustering.

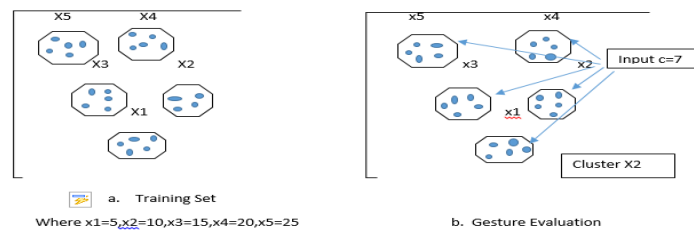


Figure 2. Clustering Analysis

Steps in LBG:

- Step1: Initialization: Set L (no of Partitions or Clusters) =1. Find the centroid for all Training data
- Step2: Splitting: Divide L into 2L separation
- Step3: Classification: Classify the set of Training data X_k into one cluster C_i according to the nearest neighbor rule
- Step4: Codebook updation: update the code word for every cluster by computing the centroid in each cluster
- Step5: Termination 1: The overall distortion D at each iterative is related to each value. If D is below threshold, then go to step 6, otherwise go to step 3
- Step6: Termination 2: If L equals vector quantization code book size, then stop, otherwise go to step 2

2.4. Speech Synthesizing

The text that is retrieved as an output of HMM is taken as an input of speech synthesizer. TTS (Text To Speech synthesizer) is the process of reading a Text/Word aloud. TTS has two blocks. User Interface and Database. User interface converts raw text into words. This process is called as Pre-processing/Normalization. The Front end, then assigns Phonetic transcription to each word and divides it into different units like Phrases/clauses/sentences. Back end often referred to a Synthesizer converts the Symbolic representation into sound by matching the Database.

Algorithm: Gesture to Sound Algorithm (GTS)

- Initially collect the gestures as input.
- Consider A_{ij} as transition states from i to j.
- b_j (Ok) is based on discrete observation of random states.
- Binarization: 0 & 1 are calculated based on
- $x(t)$: towards x axis

$y(t)$: towards y axis
 If $x(t) > 0$ and $y(t) > 0$ then,
 For ($i=1$ up to C_n) => match gesture identify C_i
 C_n : Centroid of n training set
 Training set: Comparison done in codebook database
 Choose a window function of finite length
 Place a window on top of the signal at $t=0$
 Then truncate the signal
 Compute Fourier Transform for the truncated signal
 Train the training set
 Evaluation of gestures using Viterbi algorithm is done
 Calculate Trellis (distance between ideal encoder input and actual received signal)
 Find lowest weighing path
 Add / compare / select the decision bits
 Reorder the output bits
 Text / sentence is identified and saved in a temporary memory
 Text is taken as an input to TTS
 Front end accepts the text
 while (file! =EOF)
 {read (file)}
 Separate the text into tokens
 Divide the tokens into phrases / clauses
 Finally, connect to database identifying the sound from the corresponding phrases.

2.5. Visually impaired Contact Search

First receive the gesture mode ON
 Draw pattern using GTS algorithm. Then character acknowledgement is done
 for($i=0; i<5; i++$) in screen, numbering each row with contact name
 Example:
 AAA
 AAB
 AAC
 Then speaker mode is ON
 This reads out the first 5 contact list using TTS synthesizer
 Later, gesture mode is ON
 If numbering is less than or equal to 5
 Call is placed
 Else if character is found then,
 Search continues
 Stop

3. RESULTS AND ANALYSIS

Gesture is evolved with the comparison of multiple clusters. Each character is matched with 'N' clusters. Consider, if we refer 2 characters 'a' and 'd' which matches with 5 cluster and has 5 transition states. First 3 clusters are same and the last 2 clusters differ as shown in figure 3. This type of verification is done with the help of Viterbi algorithm. Maximum characters are matched with codebook. Test set with 20 examples of sample letter shows only 5% of character mismatches. Mismatches can be reconnected by the visually impaired by re-entering the gestures. In figure 4, consider an instance as N->no. of HMM states, K->code book clusters.

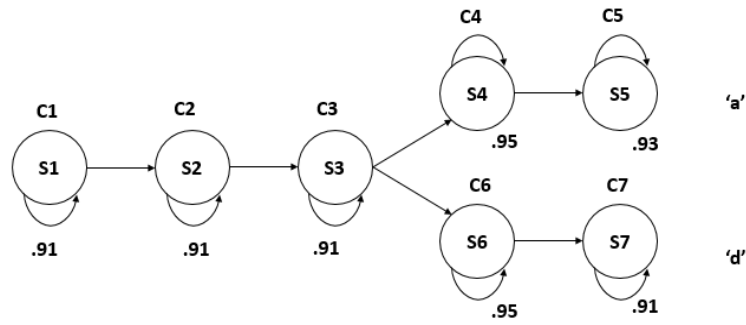


Figure 3. Transition State of 'a' and 'd'

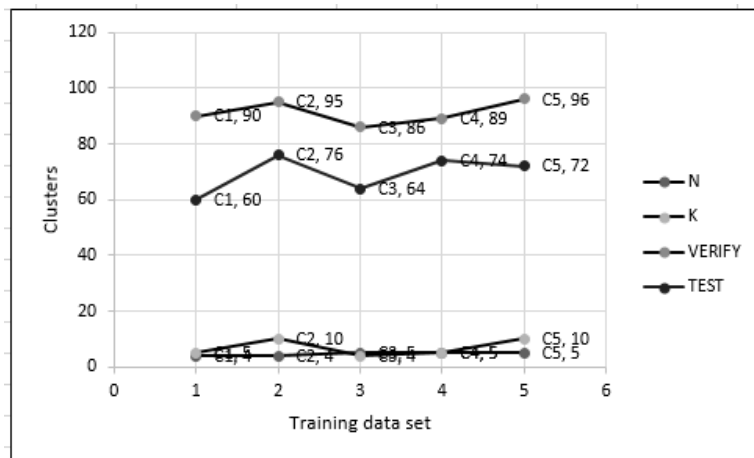


Figure 4. Training Data Set vs Clusters

TTS search: Visually impaired can go to contact list, search the specific person and place a call without a normal person help. After moving to contact list the gesture mode ON, which takes the character as an input and sound as acknowledgement. Shortlisted contacts are numbered and the first 5 contacts are confirmed through voice. If the number is pressed, call is placed to that corresponding person. If the user cannot find the contact from the shortlist he can then move to the next word. Hence, call can be placed by a visually impaired successfully.

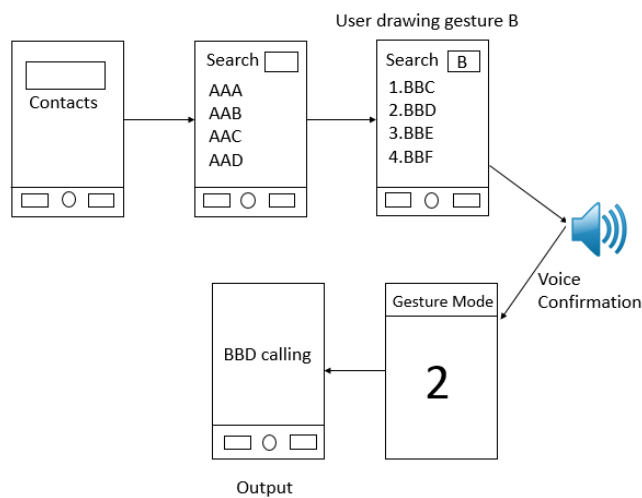


Figure 5. Retrieving Contacts for Placing Calls

4. CONCLUSION

The greater part of the existing mobile Apps have been originally created for sighted people. This paper deals around mobile apps design, concentrating mainly on amaurotic users. Speech synthesis has faraway been an imperative technology tool and its application in this field is denoting and boundless. It allows environmental obstacles to be eliminated for people with impairment. We have suggested an approach for designing, recognizing, and learning human actions by applying the Hidden Markov Model. HMM is an effectual parametric representation and is viable to characterize stochastic processes. In lieu of using geometric features, we convert the gestures into subsequent symbols. HMMs are hired to exhibit the gestures, and their parameters are studied from the training data. The gestures can be identified by assessing the trained HMMs. This application through pattern coordinating, gesture recognition and Text to speech integration would make dialing and messaging from smart phones possible and smooth for visually impaired through voice confirmation.

Our paper work deals with only English language and alphabets. It can also be used for other languages which can be implemented as a future work. Also, this paper is designed mainly for GTS conversion. The same algorithm and concept can be used for other applications like playing music, downloading app in a play store, knowing current time, etc in an Android mobile.

REFERENCES

- [1] Siddhesh Siddhesh Baravkar, Siddhesh R., Mohit R. Borde, and Mahendra K. Nivangune. "Android text messaging application for visually impaired people." *IRACST Engineering Science and Technology: an International Journal*, vol.3, no. 1, 58-61, 2013.
- [2] Sierra, Javier Sánchez, Joaquín Selva, Roca Togores, and Raylight Soluciones Tecnológicas. "Designing Mobile Apps for Visually Impaired and Blind Users Using touch screen based mobile devices: iPhone/iPad." ACHI 2012 :The Fifth International Conference on Advances in Computer-Human Interactions, pp.47-52, 2012.
- [3] Jyothi, B. Divya, and R. V. Krishnaiah. "Haptic Technology-A Sense of Touch." *International Journal of Science and Research (IJSR)*, vol. 2, no. 9, pp. 381-384, 2013
- [4] Anupam Alur, Anupam, Pratik Shrivastav, and Aditya Jumde, "Haptic technology: A comprehensive review of its applications and future prospects" *International Journal of Computer Science and Information Technologies*, vol.5, no.5 pp.6039-6043, 2014.
- [5] Pampattiwar, Sonal R., and Anil Z. Chhangani. "Smartphone Accessibility Application for Visually Impaired." *International Journal of Research in Advent Technology*, 2, no. 4, pp.377-380, 2014.
- [6] Ashraf, Anam, and Arif Raza. "Usability Issues of Smart Phone Applications: For Visually Challenged People" *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, no. 5, pp.760-767, 2014.
- [7] Sharma, Devika, and Ranju Kanwar, "Text to Speech Conversion with Language Translator under Android Environment." *International Journal of Emerging Research in Management & Technology*, vol.4, no.6, pp.96-100, 2015.
- [8] Ketaki B. Tharkude, Aishwarya K. Wayase, Pratiksha S. More, Sonali S. Kothey. Smart, Android Application for Blind People Based on Object Detection, *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no.10, pp.6019-6020, 2015.
- [9] Khan, A. and Prakash, G., Design and Implementation of Smart Glass with Voice Detection Capability to Help Visually Impaired People, *International Journal of MC Square Scientific Research*, vol.9, no.3, pp 54-59, 2017
- [10] Prakash, G., Saurav, N., & Kethu, V. R., "An Effective Undesired Content Filtration and Predictions Framework in Online Social Network", *International Journal of Advances in Signal and Image Sciences*, vol. 2, no. 2, pp. 1-8, 2016.
- [11] Olanrewaju, R. F., & Azman, A. W., "Intelligent Cooperative Adaptive Weight Ranking Policy via dynamic aging based on NB and J48 classifiers", *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, vol. 5, no. 4, pp. 357-365, 2017.
- [12] Sulthana, R., & Ramasamy, S., "Context Based Classification of Reviews Using Association Rule Mining, Fuzzy Logics and Ontology", *Bulletin of Electrical Engineering and Informatics*, vol. 6, no.3, pp. 250-255, 2017.
- [13] Rao, R. R., & Makkithaya, K., "Learning from a Class Imbalanced Public Health Dataset: a Cost-based Comparison of Classifier Performance", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 4, pp. 2215-2222, 2017.