International Journal of Advanced Science and Technology
Vol. 29, No. 06, (2020), pp. 6922-6931

Voice Based Speech Recognition for Blind Person

KattaSai Teja[1], Nalabolu Manikanta[2], K M Mehata[3], SudalaiMuthu T[4]

[1,2]Student, B. Tech-Final Year, [3]Senior Professor, [4]Associate Professor
[1,2,3,4] Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India
[1]saitejakatta281@gmail.com, [2]nalabolumanikanta@gmail.com, [3]kmmehata@gmail.com, [4]sudalaimuthut@gmail.com

Abstract

Blindness is a biggest challenge in the society. The blind people have their own challenges in identifying the neighbors. The visually challenged people are in the dire need of identifying the new people in their day to day life. Though, there are many devices are existing to help them, there is dire need for a device to help them to identify the neighbor. The android based mechanism is proposed to help the blind people in their day to day life. The aid to help visually challenged person is designed by using discourse concepts and isolating the person face through cluttered background. The smartphone is used as a tool so it reduces the external necessary devices. The discourse correspondence and speech recognition engine were utilized to identify the person. The proposed application is designed in Android platform and tested in the testing environment with 5 blind persons in their day to day activity of their living place. The results are analyzed and discussed with impact of chat bot application.

Keywords–*Speech recognition, Automated reading devices, Android App Development for Blind People, Google Voice recognition*

1. Introduction

The discourse is the source to communicate words, speech with one another. It also permits machining framework to transform the various approaching discourse signals into orders via the way toward recognizing as well as comprehension. The voice is one of the fundamental, proficient and normal type of technique for different individuals in order to have association with the each other. In current scenario, innovations of discourse are easily accessible for a particular constraint [4].

Undoubtedly, advancements of these types help in empowering machines to react accurately as well as dependably to human voices increases the easiness of needy people. It provides a very special, helpful and essential type of assistance [2]. The frameworks for interpreting, understanding and abridging universal discourse archives are complex in nature, for example, communicate news, gatherings, talks, introductions and voice messages. The Syntactic sentence structures are characterized by worldwide imperatives on word arrangements. The decision can be made based on the equations as shown in Equation 1 – 3.

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} P_{\Lambda,\Gamma}(\mathbf{W}|\mathbf{X}), \tag{1}$$

$$P_{\Lambda,\Gamma}(\mathbf{W}|\mathbf{X}) = \frac{p_{\Lambda}(\mathbf{X}|\mathbf{W})P_{\Gamma}(\mathbf{W})}{p(\mathbf{X})}, \tag{2}$$

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} p_{\Lambda}(\mathbf{X}|\mathbf{W})P_{\Gamma}(\mathbf{W}), \tag{3}$$

ISSN: 2005-4238 IJAST

The two issues are to be focused; 1). over age issue: produces right sentences as well as numerous erroneous sentences. 2).Uncertainty issue: the quantity of syntactic ambiguities in a single sentence turns out to be progressively unmanageable with the quantity of expressions [6-8]. From the scientific methodology, the objective of automatic voice recognition is to be calculated to the best possible word sequence W, given the communicated speech signal X, where optimality refers to increasing the a posteriori possibility as shown in Equation 3-4.

A portion of the significant confinements of the visual impedances face are:

- To get to composed data,
- To work gadgets with complex UIs,
- To get direction and versatility support,
- To see TV broadcastings, motion pictures, live exhibitions or shows.

The term composed data is utilized here to incorporate all print material, electronic material and written by hand material that incorporate content, logical images and graphical portrayals [9]. At the point when the composed data is language based, it is clear to express that discourse to handle it, it is the immediate other option or substitute methodology to pass on the data to a visually impaired individual. It is exceptionally useful for in-part located people. The audiotapes were generally utilized by the visually impaired network to get to data in books, magazines and papers [11-12].

$$r^* = (r+1)\frac{n_{r+1}}{n_r} \tag{4}$$

Basically, the discourse acknowledgment applications have extensively 3 classes. In disconnected word acknowledgment frame works as each word is verbally expressed with stops, so end-pointing methods can be used for recognizing the word limits dependably [13]. Secondary, profoundly obliged order along with various control applications uses the little vocabularies, constrained to specific phrases, yet utilize associated word or ceaseless discourse. Finally, enormous jargon constant discourse frameworks have vocabularies of a few but huge number of words since quite a while ago, spoken in a characteristic design.

The latter is the easiest to use yet in addition the most testing to execute [14]. The most exact discourse acknowledgment frameworks in the exploration world are still extremely moderate as well as expensive for being utilized in viable and the enormous jargon nonstop discourse applications on a wide scale as shown in Figure 1.

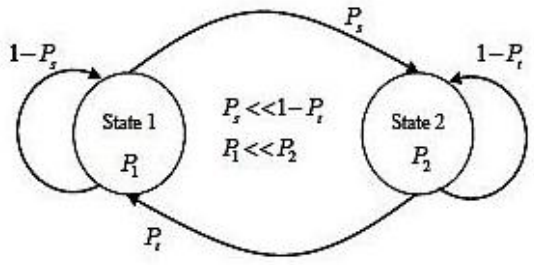


Figure 1: State transition diagram of State P1, State P2

Discourse innovation may and should utilize the predominant etymological data levels: the lexical, the syntactic, the semantic and the prosodic properties. Metalinguistic and paralinguistic data are some additional levels equipped for passing on relevant data too. Speaker attributes may likewise be utilized for this reason. Changing dialects could be thought of, if fitting, in a similar extension.

Programmed change of content into discourse or of discourse into other content is pretty much clear these days. It is being accomplished increasingly more successfully and in a reasonable manner that prompts pervasiveness. Anyway, a ton stays to be done in the upper layers of the discourse chain that interface with the data that is being passed on.

Discourse input is large hampered by acoustic obstruction and it has stayed juvenile for some applications. A couple ones, similar to machine or terminal voice orders, have been yet to get low acknowledgment from date clients. Specialized methods for conveying discourse interface arrangements is using intelligent part and a correspondence part as shown in Figure 2.

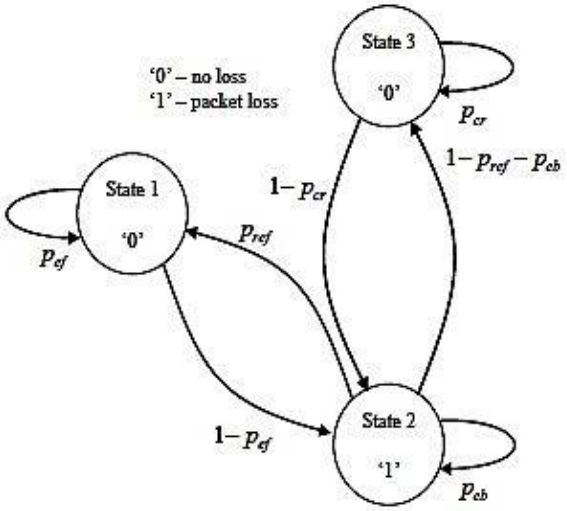


Figure 2: State transition diagram of State 1, State 2 and State 3

This technique is best for speaker subordinate application where the speaker regularly utilizes long sentences. This strategy isn't appropriate for a speaker acknowledgment that utilizations short words [15]. The last as mentioned utilizes the probability proportion approach. The probability proportion is characterized dependent on two circumstances to quantify the style of discourse spoken. The principal probability is the acoustic information that could add to the acknowledgment of the speaker and the subsequent probability is the speaker could be a fraud [16]. The standardization dependent on the back probability was additionally examined. The distinction between the standardization dependent on the proportion probability and the back probability is to guarantee that the asserted speaker is recorded for standardization. Both the strategies acquire nearly a similar degree of adequacy. The two techniques permit framework to diminish the need to rely upon just asserted speaker.

In most recent couple of years, because of nonstop gradual upgrades the discourse acknowledgment has got a solid medium to make an interpretation of verbally expressed words into writings [17]. Different virtual products like Google Voice and Siri have permitted versatile clients to simply utilize their voice to work on their mobile phones. This paper is an activity to plan a voice worked apparatus which expects to make distinctively capable of individual free, so they can act in their assessment without anyone else, without the assistance of any associate [19].

ISSN: 2005-4238 IJAST

2. Related Work

There are many related works are going on in the filed of application to the blind people.

Sandeep Musale *et al.* [18] have designed a application to find the location to travel for blind people. They have used OCR based application using MATLAB. They have converted image to text using the matlab application. They have tested the application in the defined environment. Their application yielded efficient accuracy but it invited additional over haead operation on the ground.

Deepthi Jain B *et al.* [3] have presented a system to help the visually challenges people to identify the obstacles in the travel path. They have developed the application as wearable aid. They have processed language commands. Their model was able to identify the obstacle and sign board in their day to day life. However, the identified images are required to be stored in the database in advance. It is also a drawback that they were using static database and no provision for dynamic update of the database.

Ani R *et al* [1] have introduced a combined hardware application by camera, sensors and Raspberry Pi. Their interface were able to identify the obstacles and route map for the blind people. Their application were controlling secondary units for blind people.

Kiran Rakshana R *et al.* [5]have presented a smart navguide system for visually impaired application for blind people. They have developed three module voice searching, voice processing, image processing. They have used Raspberry Pi, Camera and other input devices for capturing images. They have employed feature extraction by binarization. They were able to achieve real time cost advantage.

PankarDipak *et al.* [10] have developed intelligent cloud based system for visually impaired. They have processed the speech for identifying the needs. They have used AES technique for speech recognition. They have made provision for mobile contacts and automatic dialing. They have also facilitate tracking facility to the visually challenges people.

3. Discourse based Speech Recognition Application

Discourse is a basic method for conveying words, yet also a boundless scope of human feelings. Thusly, discourse prepared for android applications. The proposed calculation considered the outline choice techniques that measuring element wealth of a picture by removing the casings that have a higher chance of containing biased features. The figure includes lavishness, the information (distinguished face) picture can pre-handled to a standard size and changed over to grayscale. By performing face recognition and considering facial area, the other non-face substance of the casing doesn't meddle with the proposed calculation. The picture is standardized and it is utilizing its mean and fluctuation.

The discrete wavelet change (DWT) is used. The DWT estimates the coefficients of information of the picture. The DWT empower multi goals investigation of the given picture. While the main level DWT presents the coefficients for the better subtleties of the picture, the second level DWT encodes the global highlights while concentrating less on fine subtleties. The proposed model is illustrated as shown in Figure 3.

The proposed application for android Speech API gives acknowledgment control, foundation administrations, expectations and backing for numerous dialects. It provides a basic functionalities for blind people. The images are examined for identification. The four conditions set up for the examination issue establish the amalgamation of the waveform from the coefficients.

ISSN: 2005-4238 IJAST

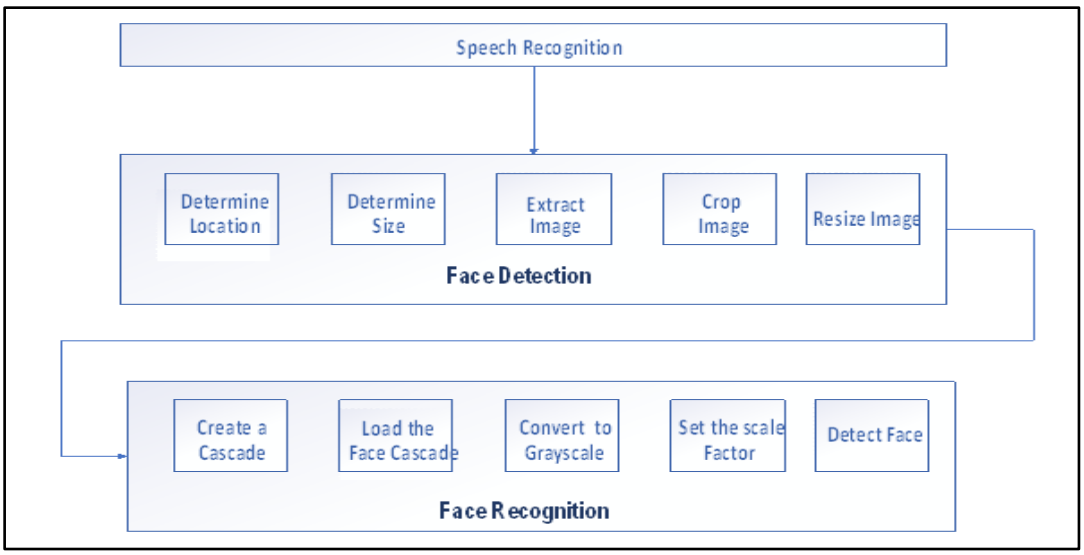


Figure 3. Architecture of Discourse based Speech Recognition

The feature extraction is carried out by the ASR feature extraction techniques as shown in Figure 4.

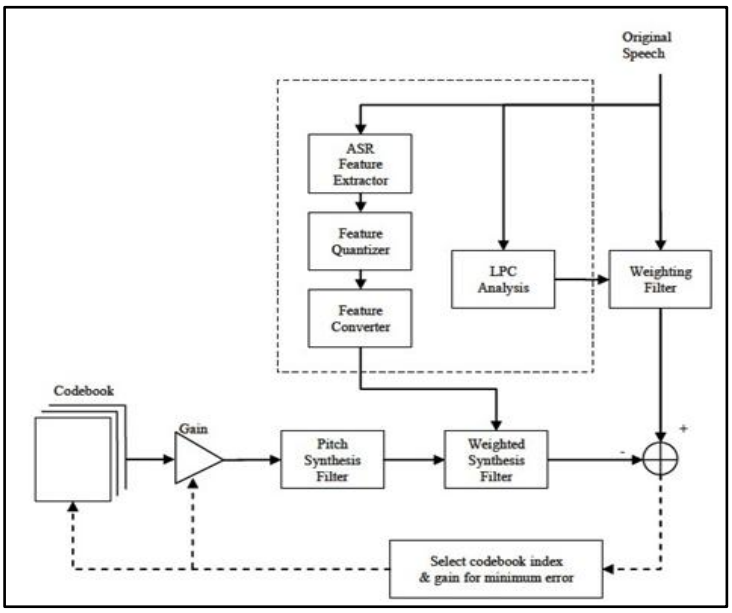


Figure 4: Feature Extraction – DWT Process

Because of the symmetry property, increasing the example estimations of a subjective waveform by the comparing test estimations of the conjugate of the perplexing sinusoid at a recurrence and adding the items must be equivalent to the coefficient of that recurrence part with a scale factor N, the quantity of tests.

Bit stream-based procedures for hearty ASR acquire discourse acknowledgment parameters straightforwardly from the bitstream transmitted to the beneficiary over computerized portable systems. The contrast between bitstream-based systems and strategies that work on the decoded discourse is that bitstream-based procedures stay away from the progression of reproducing discourse from the coded

discourse parameters. These incorporate component changes from the element portrayals utilized in the discourse coding calculation to the element portrayals utilized in ASR and systems for highlight remuneration in the bitstream space. Besides, the presence of system situated clamor sources, for example, discourse coding mutilations and channel transmission mistakes has prompted the advancement of remuneration procedures in the sign space, highlight space, and model space. A concise outline of the procedures produced for the bitstream based methodology and the system arranged clamor remuneration is given here. The installed android application has communicating with the headset and camera for the inputs as shown in Figure 5.

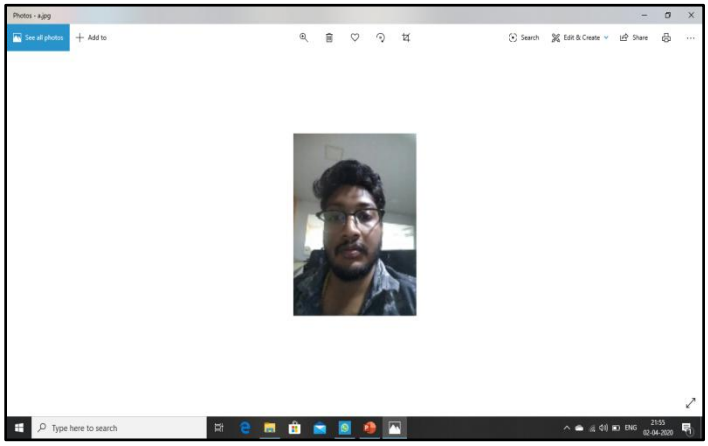


Figure 5: Image Capturing

The captured image is analyzed and identified with reference to the database. The identified image and its discourse is communicated to the blind person through headphone setup.

4. Results and Discussion

The developed android based application is tested by installing the app in five android mobiles as shown in Figure 6. The chat box application is also installed along with it as shown in Figure 7.

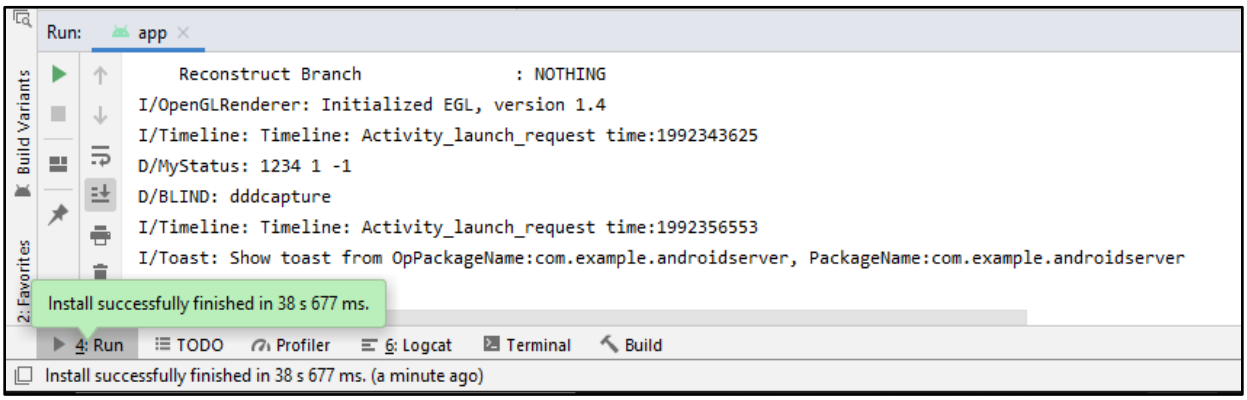


Figure 6: Installation - Proposed Application

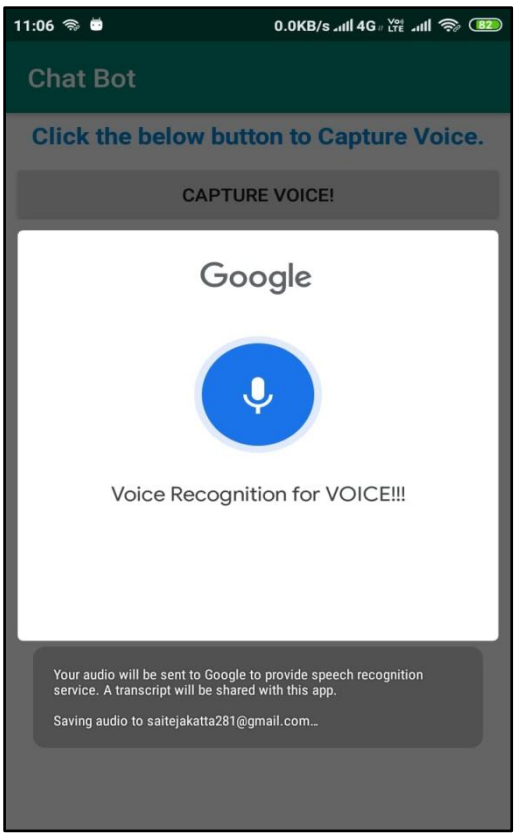


Figure 7. Installation – Chat Box Application

The application is tested with five blind people at their residence. Their accuracy of identification of person through image and voice are tested. The results are recorded and analysis. The average of accuracy is computed by the ignoring the outline data of five users. The processing time is also recorded for each identification. The average processing time and accuracy is recorded as shown in Table 1. The SNSV [5], IC_BSVI [10] have also experimented with the proposed model for validation and comparative purpose. The Accuracy and Processing time are considered as performance evaluation parameters. The accuracy is defined as ratio of True_Positive, True_Negative, False_Positive, False_Negative as shown Equation 5.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FT+FN} \quad (5)$$

The Processing time is considered for a Image Capture of 72 PPI dynamically, It can be defined as time between the image captured and the image identified as shown in Equation 6.

$$\text{Processing Time} = Time\ (Identified - Captured) \quad (6)$$

Performance Metrics	SNSV	IC_BSVI	Proposed Discourse Based Model
Accuracy in %	92.45	93.2	96.5
Processing Time in ms	1400	1600	1200

The accuracy of the algorithms is plotted as shown in Figure 8.

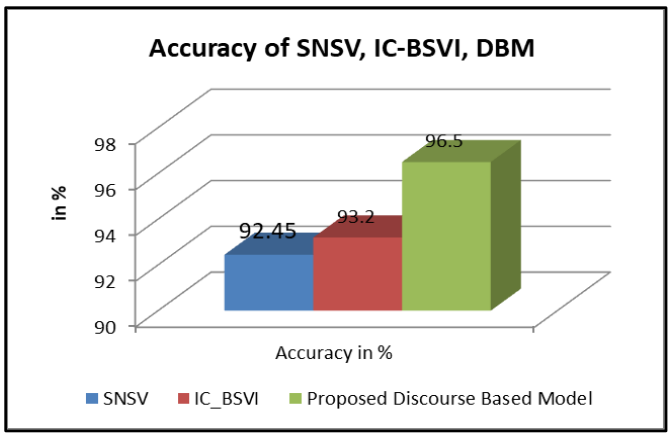
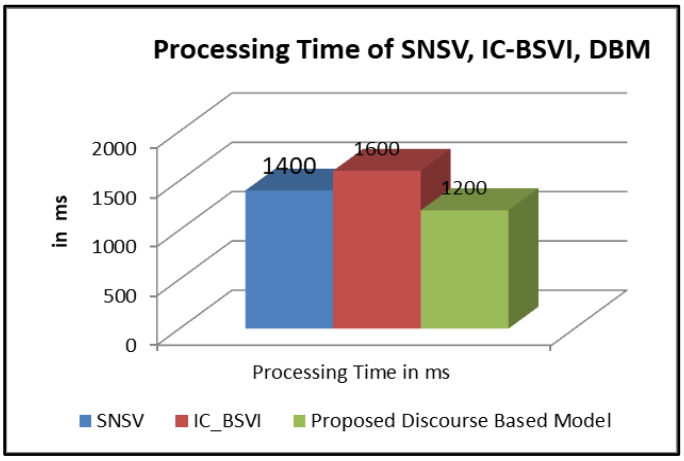


Figure 8: Accuracy of SNSV, IC-BSVI, DBM

The accuracy of SNSV, IC-BSVI and DBM algorithms are 92.45%, 93.2% and 96.5% respectively. The DBM algorithm is yielded 96.5% accuracy as outperformed as 4.3% and 3.4% than SNSV and IC_BSM algorithms respectively.

The processing time of SNSV, IC-BSVI, DBM models are plotted as shown in Figure 9.



The processing time of SNSV, IC-BSVI and DBM algorithms are 1400 ms, 1600 ms and 1200 respectively to process the 72 PPI image. The DBM algorithm is taken 1200 ms as outperformed as 14.28% and 25 % than SNSV and IC_BSM algorithms respectively.

5. Conclusion

The discourse based android application is developed for helping the blind people. The Discourse based Model is used for identifying and dialogs. The designed model is experimented with android mobile by 5 blind people in their residence area. The existing SNSV and IC_BSVI model also experimented for validation and comparison. The accuracy and processing time have been recorded. The average of accuracy and processing time plotted and analyzed. It is found that the proposed DBM outperformed in accuracy as 4.3% and 3.4% than SNSV and IC_BSM algorithms respectively. It is also found that the DBM algorithm is taken less processing time as 14.28% and 25% less than SNSV and IC_BSM algorithms respectively. The Future Work shall focus on the development of fully discourse based model to support the blind people.

ISSN: 2005-4238 IJAST

References

1. Ani R, Effy Maria, J Jameema Joyce, “Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method,” IEEE International Conference on Innovations in Green Energy and Healthcare Technologies, Nov 2017.
2. K. A. Bharadwaj, M. M. Joshi, N. S. Kumbale, N. S. Shastry, K. Panimozhi and A. Roy Choudhury, "Speech Automated Examination for Visually Impaired Students," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 378-381, doi: 10.1109/ICIMIA48430.2020.9074847.
3. Deepthi Jain B, Shwetha M Thakur and K V Suresh, “Visual Assistance for Blind using Image Processing,” IEEE International Conference on Communication and Signal Processing, April 2018.
4. S. Islam, M. J. Rahimi and K. Arham Kabir, "The Third Tab (3T): Assistant for Visually Impaired People : ATmega based automated blind cane," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934623.
5. Kiran Rakshana R, Chitra C., 2019. A Smart Navguide System for Visually Impaired, International Journal of Innovative Technology and Exploring Engineering, 8(6S3), pp. 182-184.
6. P. Lei, M. Chen and J. Wang, "Speech enhancement for in-vehicle voice control systems using wavelet analysis and blind source separation," in IET Intelligent Transport Systems, vol. 13, no. 4, pp. 693-702, 4 2019, doi: 10.1049/iet-its.2018.5094.
7. V. V. Mainkar, T. U. Bagayatkar, S. K. Shetye, H. R. Tamhankar, R. G. Jadhav and R. S. Tendolkar, "Raspberry pi based Intelligent Reader for Visually Impaired Persons," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 323-326, doi: 10.1109/ICIMIA48430.2020.9074948.
8. K. Nosirov, S. Begmatov and M. Arabboev, "Display Integrated Mobile Phone Prototype For Blind People," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-4, doi: 10.1109/ICISCT47635.2019.9011919.
9. Pandiaraj, S., Sudalai Muthu, T., Prioritization of replica for replica replacement in data grid, International Journal of Recent Technology and Engineering, 2019, Vol: 7, Issue: 5, pp. 245-248.
10. PankarDipak N., PawarPranjali N., AvhadPriyanka A., 2016, TapTapSearch : A Intelligent Cloud Based System For Visually Impaired, International Journal of Engineering Development and Research, pp. 863-865.
11. Puri, V., Ramesh Babu, A., Sudalai Muthu, T., Korabu, K., A Review of Sensor Deployment Problem in Wireless Sensor Network, Communications in Computer and Information Science, 2019, Vol: 922, pp. 259-268.
12. Puri, V., Ramesh Babu, A., Sudalai Muthu, T., Potdar, S., An Effective Optimisation Algorithm for Sensor Deployment Problem in Wireless Sensor Network, Communications in Computer and Information Science, 2019, Vol: 92, pp. 248-258.
13. Ranjana, P., Lakshmi Sridevi, S., Sudalai Muthu, T., Vikram Gnanaraj, V., Machine Learning Algorithm in Two wheelers fuel Prediction, Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019, 2019.
14. Rohini, A., Sudalai Muthu, T., A weight based scheme for improving the accuracy of relationship in social network, International Journal of Innovative Technology and Exploring Engineering, 2019, Vol: 8, Issue: 11, pp. 3040-3043.
15. Rohini, A., Sudalai Muthu, T., An Analysis of Structural Properties in Social Network Links in Educational Data Mining, 2019 International Conference on contemporary Computing and Informatics (IC3I), 2019, pp. 294-297.
16. Rohini, A., Sudalai Muthu, T., Weight based approach for improving the accuracy of relationship in social network, Journal of Advanced Research in Dynamical and Control Systems, 2019, Vol:

11, Issue: 8, pp. 188-192.
17. Rohini, A., Sudalai Muthu, T., A weight based scheme for improving the accuracy of relationship in social network, International Journal of Innovative Technology and Exploring Engineering, 2019, Vol: 8, Issue: 11, pp. 3040-3043.
18. Sandeep Musale, Vikram Ghiye, “Smart Reader For Visually Impaired,” IEEE Xplore International Conference on Inventive Systems and Control, June 2018.
19. B. Wang, W. Xiang, K. Ma, Y. Q. Mu and Z. Wu, "Design and implementation of intelligent walking stick based on OneNET Internet of things development platform," 2019 28th Wireless and Optical Communications Conference (WOCC), Beijing, China, 2019, pp. 1-5, doi: 10.1109/WOCC.2019.8770547.

ISSN: 2005-4238 IJAST