Blindophile: Mobile Assistive Gesture-Empowered Ubiquitous Input Device

Ishita Agarwal[1,2], Udit Kumar[1,3], Rachit Jain[1,4], Ruchika Chugh[1,5], and Ajith Abraham[6,7]

[1] Bharati Vidyapeeth's College of Engineering, New Delhi
[2] ishita.nishu1999@gmail.com
[3] uditkumar489@gmail.com
[4] jainrachit2011@gmail.com
[5] ruchikachugh12@gmail.com
[6] Machine Intelligence Research Labs(MIR Labs)
Auburn, Washington 98071, USA
[7] ajith.abraham@ieee.org

Abstract. Visual impairment hampers an individual's quotidian routine, diminishing the standard of life and interaction with the immediate environment. Moreover, people leveraging conventional canes face manifold challenges while commuting. This problem gets compounded when people need to access their mobile phones simultaneously. Thus, to ameliorate their experience, we excogitate a cost-effective device named, "Blindophile", meaning Blind lover, to make mobile phone accessibility effortless. Serving as a ubiquitous input device, Blindophile enables the visually impaired people to trigger the desired event on their mobile phone by performing gestures. Coupled with easy adaptability and recharging capabilities, Blindophile boasts of a hands-free mode to access mobile phones. With a paramount focus on design and accuracy, we also re-engineered a machine-learning algorithm to accomplish various functions of the mobile phone efficiently.

Keywords: Cane · Gestures · IMU · Machine Learning · Smartphone.

1 Introduction

The immanence of smartphones has changed our lives in multitudinous ways; from rudimentary applications like messaging, calling, playing music and tracking location to some convoluted functions like paying bills, booking cabs, ordering food, and much more. At present, it is almost unfathomable to survive without a smartphone. Despite the availability of a myriad of features, it is a formidable task for visually impaired people to access them with ease. Although various applications aim to alleviate the problems associated with navigation [8, 11], they do not tackle the problems of smartphone assistance. Advancement of technology in this era has not been able to fully address some of the key concerns. For instance, concurrent usage of smartphone [3] and cane becomes challenging as

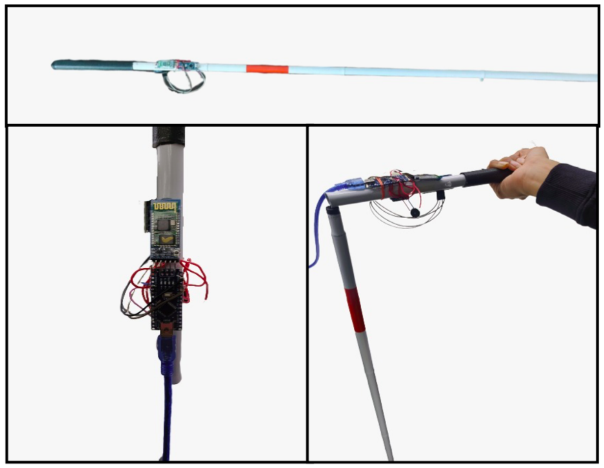


Fig. 1. Prototype of Blindophile

visually impaired people cannot interact directly with the user interfaces of mobile apps. Also, the cost associated with smartphones brings in the question of affordability. Existing solutions like virtual assistants and TalkBack applications [4] leverage speech synthesis and voice recognition for smartphone accessibility. Some of the apps like Google Assistant [10] and Siri [5] are capable of triggering certain smartphone functions through voice commands, avoiding physical interaction. However, they come with their share of the downsides. These applications fail to maintain the privacy [2] and confidentiality of the information as they require every command to be spoken out loud and clear. Moreover, the noise present in the background also interferes with these commands. According to our internal study, due to the noise in surroundings, 1 out of 5 commands are misinterpreted by these apps, thus, engendering a lack of accuracy in performing the desired action. It has also been reported that these applications are prone to miscommunication in the form of language barriers, which happens when the host applications do not support the user's vernacular. For efficient working, these virtual assistants demand uninterrupted network connection and thus, any hindrance or unavailability of the network can make them unusable.

As a result, we have presented our work on the cane itself as it is a ubiquitous device for visually impaired people. However, we have significantly transformed the usual cane to add functionalities that promise an enriching experience. Gestures, being simple and easy to perform, reflect normal human tendencies. Some of the gestures like turning, tapping, rotating, etc. can delineate corresponding actions. Blindophile allows smartphone accessibility using gestures made through the cane. It ensures that even a neophyte can learn the working without much ado and that the normal cane usage is not hampered or intervened by the specific gestures regarding smartphone accessibility.

We would like to stress upon the fact that while coming up with this lightweight, low-cost and rechargeable device, as shown in Fig. 1, design considerations were taken into account. We have employed the Inertial Measurement Unit Module (IMU), which consists of an accelerometer and gyroscope, to fetch the values of cane acceleration and gyration, produced by the gestures along all three axes. We have made use of Raspberry Pi Zero Microprocessor to implement the Machine Learning (ML) model. Since the microprocessor accepts only digital inputs, we have used an Arduino Nano microcontroller to convert analogue gestures into digital format. We have also incorporated a Bluetooth Low Energy Module (BLE) for the communication part between the microprocessor and the smartphone. In addition to this, we have used a LiPo battery as a power source to the cane. We have made sure that the components chosen are lightweight and fast in processing. Overall, the cost of Blindophile comes around 50 USD. It weighs around 350 grams and supports 10+ hours of uninterrupted use on a single charge.

Moreover, our solution supports all Android phones having Bluetooth functionality thus, making it affordable for most people. Blindophile is capable of accurately recognizing five different gestures for disparate users and setups. For the implementation of the machine learning algorithm, we have used the k-Nearest Neighbors model (kNN) [15] as the base. We have optimized this algorithm to ensure that the gestures are recognized with utmost precision and accuracy. The Machine Learning pipeline comprises:

1. An elementary methodology to glean data-set of gestures
2. Tools for training our model
3. Code to run the model on the trained data

Based on our discussions with the visually impaired, we also came up with an Android App, which triggers actions based on the gestures performed through the cane. To provide different flavours of customization, the gestures and corresponding actions are kept user-programmable. We discuss the technical aspects of Blindophile in Section 3.

2 DESIGN CONCEPTS

This segment presents the concepts augmenting our system architecture, and the corresponding challenges encountered while designing the system.

2.1 Understanding the Design Framework

Gestures provide numerous substantial advantages in comparison to other means of cane-based interaction (e.g. buttons, trackpad, and joystick [16]). Installing buttons or trackpads on a cane necessitates the user to alter the grip or the normal usage, making the task burdensome. On the other hand, gestures can be extremely intuitive. Even a beginner can learn to operate Blindophile within 5-7 minutes of training, thus ensuring a short learning curve. Besides, unlike buttons, seamless addition and reprogramming of gestures are feasible without any physical modification of the cane.

Earlier works had also proposed similar wearable technologies (e.g., smart glasses [12], smart gloves [14], smart watches [13], etc.) leveraging touch or gestures. However, the resultant devices were quite expensive and did not cater to the target audience. Moreover, it becomes onerous to employ touch-based interaction mode in addition to holding the cane and some other object. This is where the need of Blindophile arises. The cane is sufficiently capable and does not require any additional hardware.

2.2 Design Limitations and Ramifications

We discuss four key limitations that lend complexity to the design aspect. Pragmatically speaking, the device must robustly detect a decent number of gestures. It should ensure that the gestures do not intervene with the normal usage of the cane, and are recognizable for disparate users and environments. Second, a battery-driven device needs to consume low power to sustain at least a day without recharging. This limits us to using a relatively low-power microprocessor model. Third, as the cane is a ubiquitous device, it should be light-weight and portable. This requirement eliminates the possibility of directly mounting a phone or a large battery on the cane.

3 System Description

The principal objective of our system was to revamp smartphone accessibility. We effectuated this by gleaning data, consisting of disparate gestures, requisite for training and deploying the ML model to predict gestures on the cane. To this end, we designed a system encompassing the following critical components.

3.1 Components of Blindophile

Blindophile is composed of seven key components as shown in Fig. 2: 1) White Cane, 2) Inertial Measurement Unit (IMU), 3) Arduino Nano Microcontroller, 4) Bluetooth Low-Energy (BLE) module, 5) Rechargeable Battery, 6) Raspberry Pi Zero Microprocessor, and 7) On-off switch. White Cane forms the base for the other components. The IMU consists of an accelerometer and a gyroscope. The BLE module, as a communication link to the smartphone, facilitates the transmission of predicted gestures, and Arduino Nano is used as an analogue to digital converter. The power to the above elements is provided by a rechargeable battery of 1000 mAh capacity. The microprocessor runs the entire machine learning pipeline.

3.2 Gestures supported by Blindophile

Fig. 3 enlists the proposed gestures with their explanation. The gestures designed by us facilitate easy and quick learning. They allow differentiation from the

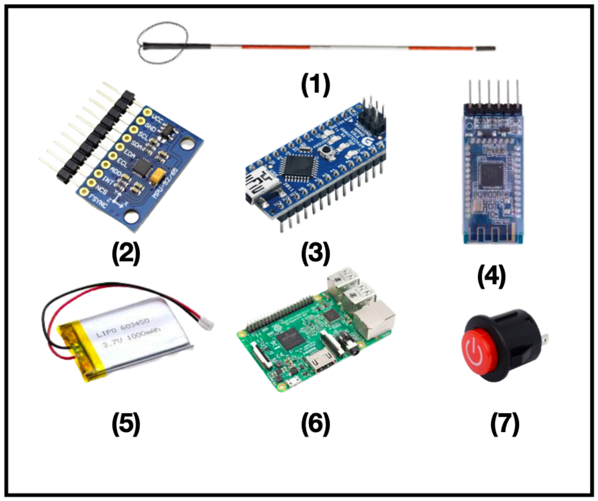


Fig. 2. Components of Blindophile

normal usage of the cane, for instance, we have used a pressure tap instead of a single tap. We have also ensured that easy improvisation of new gestures or substitution of existing ones is plausible. Moreover, if a new gesture is to be added, then the training of new data points describing that gesture can be performed quickly and simply.

3.3 Requisites for Data Preprocessing

For training, we have used 360 Pressure Taps, 240 Right Turns, 187 Left Turns, 200 Right Rotations and 200 Left Rotations in total. Our volunteers performed each gesture with the cane. Different values, labelled with the corresponding gestures, were captured by the sensor and stored in the database. Some portentous points that we considered for data collection are given below:

1. To make gesture recognition easy and accurate, variations in the orientation of the cane were included in the training set for all gestures.
2. Each gesture was approximately centred within the 1.5-seconds window because we found that all the gestures could be performed within this time, thereby marking the boundaries of each gesture.
3. To ensure that the normal usage of the cane is not hampered, we added some negative examples in our training set to compensate for the false positive results. For instance, we added single tap gestures and made sure that they are not recognized as a gesture.

3.4 ML Workflow

A machine learning pipeline was employed to understand the machine learning workflow. For Blindophile, since a stream of continuous data values is taken from

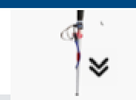
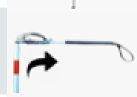
Gesture	Posture	Description
Pressure Tap (P~T)		Tapping with pressure twice
Right Turn (R~T)		Turning right by 90°
Left Turn (L~T)		Turning left by 90°
Right Rotate (R~R)		Rotating clockwise
Left Rotate (L~R)		Rotating anti-clockwise

Fig. 3. Various gestures defined in the system

the IMU module, therefore, detection of the corresponding gesture sometimes becomes difficult in real-time. Our internal studies indicate that a gesture can be completed in less than 1.5 seconds. Thus, we introduced the concept of "buffer period" which is nothing but the processor's attention time up to 1.5 seconds for each gesture. It is possible that to recognize more than one gesture, the respective buffer periods overlap and to avoid this, we fabricated a trigger mechanism that fuses a 150 ms sliding window between the overlapping buffer periods. For every 150 ms sliding window:

1. The IMU remains on standby.
2. The processor runs the ML prediction algorithm. This involves 2 steps-
   (a) Featurization: Conversion of the unprocessed/raw data obtained from the IMU into attributes/features fit for the ML model,
   (b) Application of the classification algorithm, backed by ensemble learning, to classify the categorized data into one of the gestures.
3. The recognized gesture is communicated to the smart device via the BLE module within 30 ms. Hence out of the assigned 150 ms standby time, the ML prediction must be completed in a maximum time frame of 120 ms.

The Process Flow of Blindophile is illustrated in Fig. 4. As our classification algorithm, we choose the kNN method [15]. The algorithm must work for both, gestures performed while moving or resting. To achieve this, the input of kNN i.e., the raw sensor data is carefully sampled and then converted into a feature set. This data comprises three dimensions each from the accelerometer and the gyroscope. For a buffer period of 1.5 seconds, at 100 Hz IMU sampling frequency, the raw data for each prediction instance consists of 150 (1.5 x 100) values in each of the 6 dimensions. The primary features used to draw distinctions include 5 values: count of values, range of values, rate of change of values over time, deviation span, and change in phase. Hence, in total, we have computed 30 (6 x 5) features for each training sample. After the input data is ready, we move to the implementation part. To apply kNN [15], we split the data around 70%-30% between training and testing stages and tweaked the hyper-parameters. As a result, the size of the model became 27 KB and achieved an accuracy of 94.96%.

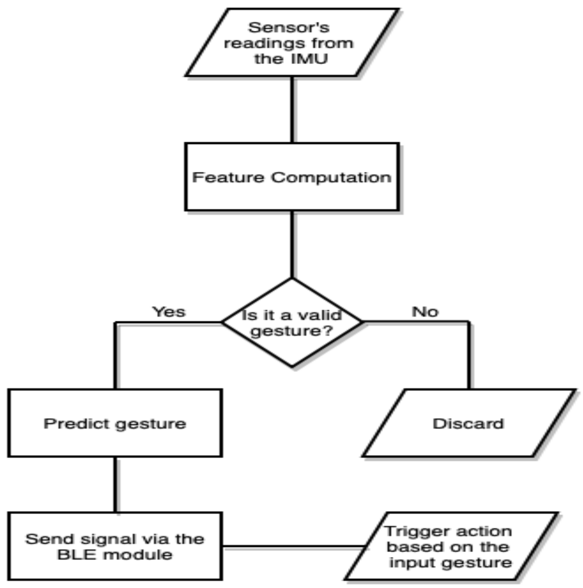


Fig. 4. Process Flow of Blindophile

3.5 Android Application

We designed an Android App intending to simplify access to the tasks like reading out notifications, time, location, and answering and declining calls, call-backs to missed calls, etc. via gestures. Initially, we configured the following gestures-

1. **Pressure Tap (P-T)**: make a call to the last person from a missed call
2. **Right Turn (R-T)**: readout the current location
3. **Left Turn (L-T)**: answer a phone call
4. **Right Rotate (R-R)**: readout the current time and the date
5. **Left Rotate (L-R)**: manage multimedia control

When the ML algorithm predicts the gesture, its corresponding hash value needs to be communicated with the connected smartphone to trigger some action. To facilitate this, we developed a typical Bluetooth terminal application capable of converting the received hash signals from the Blindophile into respective smartphone functions. This application also allows reprogramming any loaded gesture to some user-defined action as per need.

4 Evaluation

To assess the performance of Blindophile, we collected 90 instances for each of the 5 gestures from 15 participants. We measured accuracy using: 1) precision, the frequency of correctly recognised instances of the gestures 2) recall, the frequency of the times when the gesture performed was recognised correctly. The below-given confusion matrix, in the form of the heatmap (Fig. 5), presents our results. In the table, each cell value depicts the frequency for a gesture in the nth row being recognised as a gesture in the mth column. We reached the

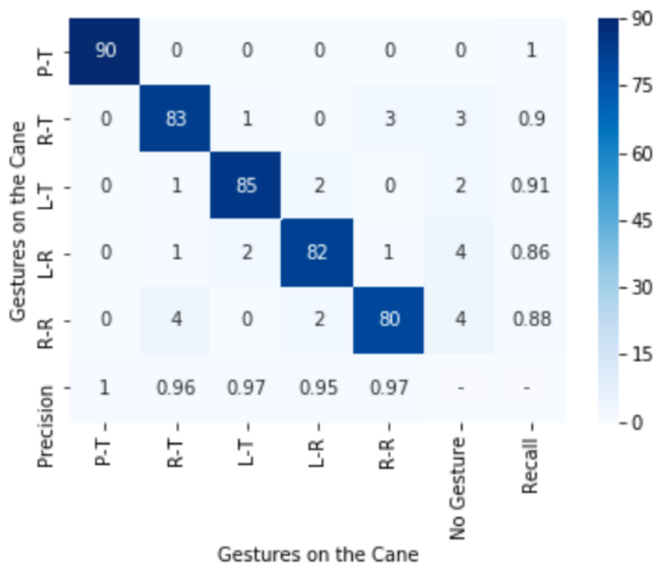


Fig. 5. Confusion matrix for gesture recognition using kNN

following conclusions-

First, Blindophile recognizes pressure-tap (P-T) with a precision value of 1 (all 90 instances are recognised correctly). Second, our model sporadically mispredicts a left rotation (L-R) as a left turn (L-T) or a right rotation (R-R). Despite a relatively low precision value, left rotate (L-R) still achieves a recall value of 86%. Overall, Blindophile achieves a precision of 0.97 and a recall of 0.91 across all the gestures.

A high recall value is a yardstick for low false negatives whereas a high precision value is the same for low false positives. We drew insights into the number of false positives and false negatives for each gesture (Fig. 6). Further, simple computation yielded the false acceptance rate (FAR) and false rejection rate (FRR). FAR indicates the probability of an incorrect gesture being recognized as correct. Similarly, FRR indicates the probability of a correct gesture being rejected as incorrect. Overall, across all instances of the gestures (90 x 5), the FAR turned out to be 3.77% and the FRR stood at 6.67%. These results underscored the robustness and efficiency of Blindophile.

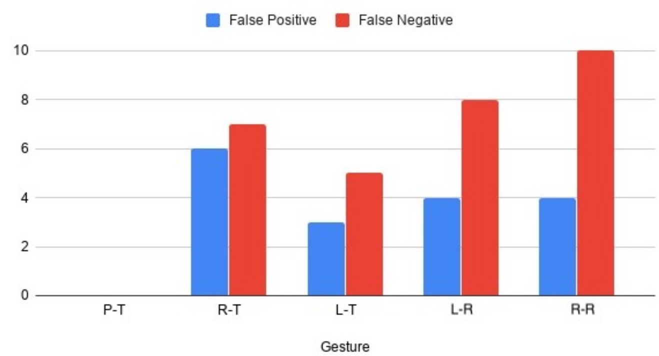


Fig. 6. Number of false positives and false negatives for respective gestures

Moving forward, as we understand, with more training data available for a gesture, the prediction accuracy naturally increases. On the flip side, as the number of target gestures increases, the accuracy decreases. We assumed k as the number of target gestures and m as the number of training examples per gesture. As a result, in our model, k varied from 1-5 and m varied from 1-4. Keeping a prime focus on the extensibility of the model, we analysed the relationship between accuracy and the number of target gestures, k. We illustrated our findings in light of Fig. 7.

Overall, our algorithm was efficient and achieved an average accuracy of 97.5% for k=1 and 81% for k=5. Essentially, the findings corroborated our hypotheses. Initially, the accuracy decreased as the number of target gestures was increased. However, as evident from the graph, the accuracy soon stabilized. This led us to conclude that the number of target gestures can be increased, as per the requirement, without any significant impact on the performance after a certain point. Our results also demonstrated that Blindophile can learn effectively from

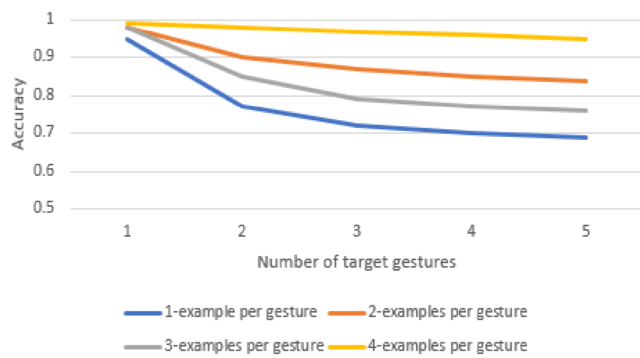


Fig. 7. Variation of accuracy with the number of target gestures

copious data. Additionally, with an increased number of training examples, we surmise improvement in the recognition rate as well.

5 Discussions with Related Works

Visually impaired people also intend to make use of smartphones in manifold ways [1]. However, their disability acts as an impediment, thus making the above mentioned task daunting [3]. Nonetheless, solutions like TalkBack features on the phone [4], VoiceOver applications, and mobile applications [11, 8] have opened a vast array of possibilities. At the same time, considerable progress in the field of Artificial Intelligence has led to the advent of Google Assistant [10] and Siri [5]. Additionally, special smartphones [9] have been dedicatedly built for visually impaired people. However, they have not been able to garner much support due to limited availability and servicing. Recently in 2019, Samsung has also come up with a communication app, named "Good Vibes" [7] for the deaf-blind. With advancements in gesture recognition technology, multi-modal keypads [6] came into being, but they had to be used independently and thus, had limited usage. Moreover, they were expensive and required extensive maintenance and care; a replacement had to be done if there was any problem. Furthermore, to reduce the need for additional hardware, the concept of a smart cane came into the picture. Henceforth, a lot of work was carried out to enhance obstacle detection [17] and navigation capabilities using ultrasonic sensors, radars, etc. In this paper, our main focus area has been to ease smartphone accessibility. Our solution builds upon the previous works to develop a mobile assistive gesture-empowered interactive cane, sporting reprogrammable features. The solution proposed by us is complementary in nature as it can leverage the above-mentioned technologies in concomitance.

6 Conclusion

In summary, we introduced a device named Blindophile, a gesture-empowered user-friendly ubiquitous cane that assists visually impaired people to use smartphones efficiently. Blindophile, being a low-cost, light-weight, and portable device, sports a contactless mode for smartphone accessibility and does not require additional hardware. Our internal studies indicate that Blindophile is robust in recognizing gestures with high accuracy for a diverse number of users and environments. Coupled with an Android app, it allows reprogramming of gestures without physical modification of the cane. Additionally, a short learning curve facilitates quick adaptability by the users. Based on feedback from the visually impaired, the gesture-based cane emerges as a propitious mode for smartphone interaction and accessibility. Overall, the cost of Blindophile comes around 50 USD. It weighs around 350 grams and supports 10+ hours of uninterrupted use on a single charge..

References

1. Watanabe, T., Miyagi, M., Minatani, K., Nagaoka, H. (2008, July). A survey on the use of mobile phones by visually impaired persons in Japan. In International Conference on Computers for Handicapped Persons (pp. 1081-1084). Springer, Berlin, Heidelberg. Chicago

2. Ye, H., Malu, M., Oh, U., Findlater, L. (2014, April). Current and future mobile and wearable device use by people with visual impairments. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3123-3132).

3. Rodrigues, A., Nicolau, H., Montague, K., Guerreiro, J., Guerreiro, T. (2019). Open Challenges of Blind People using Smartphones. arXiv preprint arXiv:1909.09078. Chicago

4. Rodrigues, A., Montague, K., Nicolau, H., Guerreiro, T. (2015, October). Getting smartphones to talkback: Understanding the smartphone adoption process of blind users. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers Accessibility (pp. 23-32).

5. Abdolrahmani, A., Kuber, R., Branham, S. M. (2018, October). " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 249-258).

6. Buzzi, M. C., Buzzi, M., Leporini, B., Trujillo, A. (2014, October). Designing a text entry multimodal keypad for blind users of touchscreen mobile phones. In Proceedings of the 16th international ACM SIGACCESS conference on Computers accessibility (pp. 131-136).

7. https://www.livemint.com/mint-lounge/features/how-accessibility-apps-are-connecting-the-dots-1568967806008.html

8. Banf, M., Mikalay, R., Watzke, B., Blanz, V. (2016). PictureSensation–a mobile application to help the blind explore the visual world through touch and sound. Journal of rehabilitation and assistive technologies engineering, 3, 2055668316674582.

9. Pal, J., Viswanathan, A., Song, J. H. (2016, December). Smartphone adoption drivers and challenges in urban living: Cases from Seoul and Bangalore. In Proceedings of the 8th Indian Conference on Human Computer Interaction (pp. 24-34).

10. Ms. Kavya. S, Ms. Swathi, Mrs. Mimitha Shetty, 2019, Assistance System for Visually Impaired using AI, International Journal of Engineering Research Technology (IJERT) Rtesit – 2019 (Volume 7 – Issue 08),

11. Avila, M., Wolf, K., Brock, A., Henze, N. (2016, June). Remote assistance for blind users in daily life: A survey about be my eyes. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments (pp. 1-2). Chicago

12. Hassan, E. A., Tang, T. B. (2016, July). Smart glasses for the visually impaired people. In International Conference on Computers Helping People with Special Needs (pp. 579-582). Springer, Cham.

13. Neto, L. D. S. B., Maike, V. R. M. L., Koch, F. L., Baranauskas, M. C. C., de Rezende Rocha, A., Goldenstein, S. K. (2015, April). A wearable face recognition system built into a smartwatch and the blind and low vision users. In International Conference on Enterprise Information Systems (pp. 515-528). Springer, Cham.

14. Linn, T., Jwaid, A., Clark, S. (2017, July). Smart glove for visually impaired. In 2017 Computing Conference (pp. 1323-1329). IEEE.

15. Zhang, S., Li, X., Zong, M., Zhu, X., Cheng, D. (2017). Learning k for knn classification. ACM Transactions on Intelligent Systems and Technology (TIST), 8(3), 1-19.

16. Wahab, M. H. A., Talib, A. A., Kadir, H. A., Johari, A., Noraziah, A., Sidek, R. M., Mutalib, A. A. (2011). Smart cane: Assistive cane for visually-impaired people. arXiv preprint arXiv:1110.5156.

17. Utaminingrum, F., Sari, Y. A., Somawirata, I. K. (2019, July). Obstacle Detection for Assisting Navigation of Visually Impaired People Based on Segmentation Process. In Proceedings of the 2019 4th International Conference on Robotics, Control and Automation (pp. 27-32).