

Literature survey

Team id	PNT2022TMID51506
Project name	Real time communication system powered by AI for specially abled
Team Lead	ALBIN AJITH S
Team member 1	SASIKUMAR V
Team member 2	SASIKUMAR V
Team member 3	HEMALATHA S

Abstract:

Communication plays a significant role in making the world a better place. Communication creates bonding and relations among the people, whether persona, social, or political views. Most people communicate efficiently without any issues, but many cannot due to disability. They cannot hear or speak, which makes Earth a problematic place to live for them. Even simple basic tasks become difficult for them. Disability is an emotive human condition. It limits the individual to a certain level of performance. Being deaf and dumb pushes the subject to oblivion, highly introverted. In a world of inequality, this society needs empowerment. Harnessing technology to improve their welfare is necessary. In a tech era, no one should be limited due to his or her inability. The application of technology should create a platform or a world of equality despite the natural state of humans.

On the other hand, technology is the most innovative thing on Earth for every time the clock ticks, researchers, software engineers, programmers, and information technology specialists are always coming up with bright ideas to provide convenience to everyone. This paper shows how artificial intelligence is being used to help people who are unable to do what most people do in their everyday lives. Aligned with communication, D-talk is a system that allows people who are unable to talk and hear be fully understood and for them to learn their language easier and also for the people that would interact and communicate with them. This system provides detailed hand gestures that show the interpretation at the bottom so that everyone can understand them. This research allows the readers to learn the

system and what it can do to people who are struggling with what they are not capable of and will provide the technical terms on how the system works

Key words: Machine learning, disability application, sign language, image processing, neural networks, artificial intelligence

INTRODUCTION:

Communication should be universal without any barriers or limitations. This paper establishes a method for providing equality, turning the disabilities of the hearing and, or speech impaired individuals to abilities, creating a base where both the disabled and the able can communicate without any barrier. Our objective is to blend deaf and dumb within society and make them able to use their personal computers more effectively and efficiently. Our idea is to create sign assistance, like many applications which is using voice assistance such as Siri on iOS and Cortana on windows. There is need to develop an application that will create an interactive platform where the sign language can be translated to voice output and writing, and voice and writing input can also be converted to sign language. The bigger picture is creating an interactive model of communication for deaf and dumb people.

Developing an app will support this vulnerable society of impaired people and enhance communication among people. The application will allow ease in communication, improving their interaction, and hence better life. This project will be a noble cause and translating the sign language into understandable words is the goal. Microsoft Windows will come in handy to enhance the actualization of this application. According to the World Health Organization, the world population experiencing hearing and speech challenges approximates over 466 million people globally [1,2]. With such disability, instead unequally distributed resources, these people are vulnerable to discrimination [3]. The fact that every human being, abled or disabled, is entitled to a good life with equal opportunities calls for affirmative action [4,10]. This society requires attention from all quarters, especially on technological enhancement, to ensure the disabled get a comfortable life [5,16,20]. With the number increasing significantly, something needs to be done. The deaf and dumb are introverts, remaining

engraved in their thoughtful world. Communication, which is essential in human life, is challenging. Humans are social beings, and effective communication is necessary [6, 22,24]. The development of technology should, therefore, serve to improve their lives as well [7]. The introduction of an application that can be used by the deaf and dumb will be a great innovation. It will not only make life easier but will as well increase their life opportunities, including employability. The deaf and dumb category must be involved within technology on PC experience as they involved in technology on smartphones.

THEORETICAL BACKGROUND:

This section provides a theoretical background for D-talk to have a better understanding of the process to be used in the application [13,30, 32]. This application can provide a helpful tool for communication between the deaf and the external world. The sign language recognition program, which is required to understand sign languages, has been studied extensively for years [25,27]. The studies are focused on various input sensors, gesture segmentation, feature extraction, and classification methods. This paper is aimed at evaluating and comparing the methods used in the sign recognition systems, classification methods used and identifies the most promising approach for this project. Despite recent advances in classification methods, many of the recent works proposed to apply primarily to classification methods, such as deep learning [23, 34, 40, 42]. This paper focuses on the classification methods used in the prior recognition scheme for sign Recognition. Based on our research, HMM-based methods, including its modifications, have been thoroughly discussed in prior studies.

Machine Learning:

It is essential to choose the right strategy; Machine-learning techniques are often used to do this. Machine learning is part of artificial intelligence (AI). It can be defined as an algorithm that focuses on computer program development [14, 15, 35, 43]. Machine learning considered an application

that use to increase computer ability to learn from previous experiences [16, 17, 44]. There are three types of ML: •Supervised: The application has previous experience of new data using labeled examples to predict future events [18, 19]. •Semi-supervised: Data can be labeled and unlabeled for training [29, 36]. •Unsupervised: Data that unlabeled from experience. The system cannot figure out the right input so that the result will be wrong [20, 21, 37].

Deep Learning:

Compared to standard algorithms, neural networks can solve somewhat complicated issues at a much easier level about the complexity of algorithms. Neural networks can solve somewhat complicated issues at a much easier level concerning the complexity of algorithms [26, 30]. The neural network builds to mimic human brain neural function but with the mathematical functions [31, 33,38]. One of the neural networks is the multilayer network as shown in figure 1. It includes three layers, the input layer, many hidden layers, and the output layer [21,39]. The input layer passes data without modification. Hidden layers process the data, and the output layer converts hidden layers to output as a classification. Collecting datasets for training takes time to process [41, 45]. As the number of configuration increases, training samples increase. Most data in the world not uniformly distributed. An image recognition model using neural network.

Decision Tree:

A decision tree uses datasets that include attribute vectors that contain a collection of classification class [16, 20]. This technique builds the dataset on attributes that split the data to different existing classes until the criterion is reached [20, 31]. The decision tree has an excellent representation form to have an idea of data quickly due to easy visualization format. The structure of the decision tree contains nodes that separated into root node, inner node, and end node 'leaf'. The root node is the start of the decision has no incoming edges [32, 40]. Inner nodes include one incoming edge and at least more than one outgoing edge. A decision tree is a tool that uses most of the

time to support decisions because it shows the rule of sequential decisions [16]. Marketing people can use it to predict customers purchased by their age; for example, if the consumer age is between 30 and 45, they prefer to rent Kia cars. End nodes will get the answer as an output as a decision to the forecasting [18]

Object Detection:

For this application, object detection will be involved. It is one of the classical problems of computer vision because it depends on different elements, such as the level of lighting and the position of the user. It is difficult to have an idea where the object is located and how the image is segmented. We need to know what the shapes of the fingers are, where they are located. The software will use a saliency map to guide for locating the object. The location and size will be defined by a bounding box, rectangle shape, which located on corner coordinates. shows the example of object detection. Example of Object Detection The webcam will try to capture the sign that the user will gesture by size and location. Then, the software will use algorithms above and get the right output to depend on the input that will be produced by the user. Moreover, several conditions can help the application to give the best result, such as lightning, positioning, mirroring, and scale. Lightning and shadows condition could affect image recognition, but the system must be able to recognize the objects. Object position can be changed, the application must be able to identify the object position. The object recognition system must recognize mirrored images.

Image Recognition Process:

The image recognition process is a process that enables the input of the sign language into the application for necessary processing [20,31,46]. The process requires a sign to be made in front of the webcam. The computer captures the sign made via the webcam and stores the different images made. Images that come from the camera will be resized, and the resolution will change. The colors will change to grayscale image and then to black and white images while editing the images [25, 33,47]. There several techniques used to extract the image, such as SIFT, SURF, BRISK, and HSV algorithms. Scale Invariant Feature Transform "SIFT" is used to extract

features vectors that define local patches of the image [35]. These features are invariant to translation, rotation, and illumination, not only vectors scale in variant. There are several advantages to the SIFT descriptor [36]. For example, this descriptor is accurate than any other descriptor. It can describe the key points in the image for any object. It allows an individual feature to define the correct match with the best probability in a huge database of features [37]. Also, it can cover the full range scales and location of the image, and it is close to real-time performance.



METHODOLOGY:

This paper presents an understanding of complex hand movements. A framework based on Hidden Markov Models (HMMs) [11] is provided for modeling and recognition of complex gesture trajectories. A detection algorithm is used to detect the hand of the user, and a contour-based hand tracker is developed, which combines condensation and partitioned sampling. The proposed approach can attain automated online identification of hand movements and can effectively reject atypical movements. The experimental results indicate that the proposed algorithm can produce better results for recognition than the conventional method of hand recognition [12]. The hand gesture recognition system consists of three major parts: palm detection, hand tracking, and trajectory recognition. Figure 5 provides an overview of the hand gesture recognition process. The hand tracking function is enabled when the device senses an open hand in front of the

camera; when the user finishes the gesture, the hand gesture classification based on HMM is disabled. The basic algorithmic structure for our process of recognition is the following: -

- Detect the palm from the video and initialize the tracker with the template of hand shape.
- Track the hand motion using a contour-based tracker and record the trajectory of the palm center.
- Classify the gesture using HMM, which gives the maximum probability of occurrence of observation sequence

The research aims to provide detection, initially for four unique signs. A dataset of images was made for four unique signs to start the detection process. The dataset contained at least 1000 images. The training images will be for men and women's hands to make the application suitable for both genders. Each sign will have 50 images with different positions and light levels. After the image is taken from the webcam, the application will process the image and compare it with the 1000 images that are in the dataset. The nearest image will be taken as a decision to produce the task. The webcam may take images in the dataset with the distance between the hand and the cam around 3-4 feet.

Training Phase:

The training phase was based on storing the images in the database. The database contained images of hands, both men and women. The training was based on identifying all possible signs that can be made using one hand. For this purpose, 30 different images with different levels of lights and duration were captured and stored in the database. These images were used as training images that will help in making the right decision for the tasks. The database contained over 1000 images of unique hands and signs.

Gesture recognition:

We explored one way to identify simple hand gestures and implement two basic gesture controls: movement of the cursor and mouse click. The figure 6 describes the basic process of hand gesture recognition.

- By using vision-based recognition, the computer captures the sign to find the gesture acquisition.

IMPLEMENTATION AND RESULTS:

In the implementation phase, developers change several tasks that they were planned to do. They notice that they can build the system without preparing any training and testing images as they were plan. The code is depending on skin color and contour to find the right sign. Moreover, developers narrow the tasks to only one task which is browse websites only. Moreover, the result was precise and accurate aligned with the methodology and testing that was used. This signifies that developing modern technology assists disabled individuals specifically deaf-dumb on interacting among people. The measurement variables along with the supporting evidence from the methodology concluded that the measures taken to evaluate this study were supported all throughout. Meanwhile, the efficiency and effectiveness of the system provide the utmost benefit of disabled individuals by offering convenience and being able to make their lives easier and better for there are no required training or specificities for them to use the system. Thus, as a result, D-talk allows everyone to determine the hand gestures that are being projected and be able to come up with interpretations on enabled individuals. Hence, communications between deaf-dumb and enabled individuals are way easier and lacks misunderstandings are being prevented this time. This application can catch finger shapes by using the code for Extract skin color and draw lines around the hand. As a result, the system will recognize any element in the frame. The application main screen is shown in figure 9. Thus, users must be careful about what is inside the frame to avoid any other unwanted requests. This system will recognize any element in the box, and the brightness does not matter. D-talk is a dynamic system that includes three gestures in total to browse websites. All that users need to implement this system is WiFi connection and webcam to capture user gestures. The

following figures 10 and 11 show the hand gestures that are used for orders that the system can recognize to browse websites. It could be used as a guide for users.

REFERENCES

1. Abhishek, K. S., Qubeley, L. C. F., & Ho, D. (2016, August). Glovebased hand gesture recognition sign language translator using capacitive touch sensor. In 2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC) (pp. 334-337). IEEE. <https://doi.org/10.1109/EDSSC.2016.7785276>
2. Ahmed, M. A., Zaidan, B. B., Zaidan, A. A., Salih, M. M., & Lakulu, M. M. B. (2018). A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017. *Sensors*, 18(7), 2208. <https://doi.org/10.3390/s18072208>
3. Ahmed, M., Idrees, M., ul Abideen, Z., Mumtaz, R., & Khalique, S. (2016, July). Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2. In 2016 SAI Computing Conference (SAI) (pp. 330-335). IEEE. <https://doi.org/10.1109/SAI.2016.7556002>
4. Artemov, M., Voronov, V., Voronova, L., Goncharenko, A., & Usachev, V. (2019). Subsystem for Simple Dynamic Gesture Recognition Using 3DCNNLSTM. In Conference of Open Innovations Association, FRUCT (No. 24, pp. 571-577). FRUCT Oy.
5. Almisreb, A. A., Jamil, N., Norzeli, S. M., & Din, N. M. (2020). Deep transfer learning for ear recognition: A comparative study. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1.1 Special Issue), 490-495.[80].
6. Anderson, R., Wiryana, F., Ariesta, M. C., & Kusuma, G. P. (2017). Sign language recognition application systems for deaf-mute people: A review based on input-process-output. *Procedia computer science*, 116, 441-448.
7. Ansari, Z. A., & Harit, G. (2016). Nearest neighbour classification of Indian sign language gestures using kinect camera. *Sadhana*, 41(2), 161-182.
8. Aujeszky, T., & Eid, M. (2016). A gesture recognition architecture for

Arabic sign language communication Bayan Mohammed Saleh et al.,
International Journal of Advanced Trends in Computer Science and
Engineering, 9(4), July – August 2020, 4374 – 4382 4381 system.
Multimedia Tools and Applications, 75(14), 8493-8511.

9. Badhe, P. C., & Kulkarni, V. (2015, November). Indian sign language
translator using gesture recognition algorithm. In 2015 IEEE International
Conference on Computer Graphics, Vision and Information Security
(CGVIS) (pp. 195-200). IEEE.

10. Bukhari, J., Rehman, M., Malik, S. I., Kamboh, A. M., & Salman, A.
(2015). American sign language translation through sensory glove;
signspeak. International Journal of u-and e-Service, Science and
Technology, 8(1), 131-142. <https://doi.org/10.1109/CGVIS.2015.7449921>