Hand Gesture Recognition System

Mayank Roy
Greater Noida, India
mayank212001@gmail.com

Kadambani Sharma
Greater Noida, India
skadambani1@gmail.com

Dr. Anu Bala
Assistant Professor
anu.bala@galgotiasuniversity.edu.in

***Abstract*-Around 2.78% of people in our country are unable to speak, either because they are deaf or dumb. They rely on hand gestures and motions to communicate with others, using a form of communication known as sign language. However, it can be challenging for them to convey their messages to non-sign language users, who are not trained in interpreting sign language.**

To address this problem, one solution is to convert sign language into text, allowing others to read and understand the message. There are two major techniques available for detecting hand motion or gesture: vision-based and non-vision-based techniques. In vision-based techniques, a camera is used for gesture detection, while non-vision-based techniques involve the use of sensors.

This paper uses a vision-based technique. It consists of four modules: pre-processing, hand segmentation, feature extraction, sign recognition, and sign-to-text. In this sign language recognition paper, a sign detector was modified to detect different hand gestures and provide the corresponding gesture name from the model's database. The model was also tested for accuracy on additional gestures.

This paper is flexible and can be trained according to the user's needs. Additionally,it is portable and can be used on a laptop without any additional hardware devices.

Keywords:- Sign Language Recognition System, Open CV, Pixel First Algorithm, Landmarks First Algorithm.

I. INTRODUCTION

In coordination with the existing technologies like KNN and CNN we enhanced the model with 2 new algorithms like Pixel's first algorithm and Landmark's first algorithm.

Hand gesture recognition is an active field of research in the Human-Computer Interaction technology domain, with numerous applications in virtual environment control, sign language translation, robot control, and music creation.

1) In this paper, we implemented a hand gesture recognizer using OpenCV and Python, leveraging the MediaPipe and Tensorflow frameworks for gesture detection and recognition, respectively. We explored the fundamentals of Neural Networks, file handling, and common image processing techniques. Our approach involved two different methods: the pixel-first approach and the landmark-first approach.

2) The pixel-first approach involved capturing frames from the camera, which were then interacted with in Python using the OpenCV library. We utilized a deep neural network learning method to handle the system, as we extracted RGB matrices that represent all image frames in computers as matrices. We passed this information to a Convolutional Neural Network (CNN) that was optimized for image processing and classification, resulting in the pixel-first approach.

3) On the other hand, the landmark-first approach involved extracting hand landmarks from frames passed through a control network, rather than providing definitive answers on the hand gesture's meaning. We then passed these landmarks through a simpler neural network

architecture that provided the final classification of the hand gesture.

4) Our approaches offered various advantages over traditional methods such as the Convolutional or K-Nearest Neighbor Networks. Our models were trained using RGB matrices, negating the need for background elimination, and enabling us to build our own database for each gesture, which we trained at least 30-40 times to create an accurate and reliable model. Moreover, our methodology allowed us to train the model for different orientations of each gesture rather than collecting numerous variations with different backgrounds, thus simplifying the training process.

Finally, we incorporated American Sign Language (ASL) into our model, enabling it to recognize all 26 letters of the alphabet through hand gestures.

II. LITERATURE SURVEY

In literature review, we studied existing papers and papers related to this topic and tried to understand the existing system behaviors.

Kazuhito Takahashi[1] has developed a system using mediapipe, openCV and various deep learning algorithm such as Keras Neural Network and Pixels first algorithm and landmark first algorithm that have helped to create such a reliable system for hand detection that can be reliable and works flawlessly.

Rupesh Prajapati, et al.[2] have developed a system that can convert a hand gesture into a speech by using KNN and PCA algorithm.

Raimundo F. Pinto , et al. [3] have proposed a sign recognition system based on CNN, ANN, Contour Extraction and polygon approximation for extraction of a feature and the image processing part. There are some limitations to the system i.e dealing with the image background.

Anchal Sood and Anju Mishra [4] have proposed a hand gesture sign recognition system based on Harris algorithm for extraction of feature in which after the image pre-processing part, the feature is extracted and stored in the N*2 matrix. This matrix is further used to match the image from the database. There are some limitations to the system. The very light brown to somewhat dark brown background gives error as they are considered in the range value for skin segmentation. But the results are efficient.

Karray F., et al. [5] 2008 Human- Computer Interaction: Overview on State of the Art Int. J. Smart Sens. Intell. Syst. 1 137-159. The intention of this paper is to provide an overview on the subject of Human-Computer Interaction. The overview includes the basic definitions and terminology, a survey of existing technologies and recent advances in the field, common architectures used in the design of HCI systems which includes unimodal and multi-modal configurations, and finally the applications of HCI. This paper also offers a comprehensive number of references for each concept, method, and application in the HCI.

Murthy G. et al.[6] 2009 A review of vision based hand gestures recognition Int. J. Inf. Technol. Knowl. Manag. 2 405-410. With the ever-increasing diffusion of computers into the society, it is widely believed that present popular mode of interactions with computers (mouse and keyboard) will become a bottleneck in the effective utilization of information flow between the computers and the human.

Garg P., et al. [7] 2009 Vision based hand gesture recognition World Acad. Sci. Eng. Technol. 972-977. With the development of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. Due to the limitation of these devices the usable command set is also limited. Direct use of hands as an input device is an attractive method for providing natural Human Computer Interaction which has evolved from text- based interfaces through 2D graphical-based interfaces, multimedia-supported interfaces, to fully fledged multi-participant Virtual Environment (VE) systems.

Yeo H.-S., et al. [8] 2013 Hand tracking and gesture recognition system for human- computer interaction using low-cost hardware Multimed.

Tools Appl. Abstract Human-Computer Interaction (HCI) exists ubiquitously in our daily lives. It is usually achieved by using a physical controller such as a mouse, keyboard or touch screen. It hinders Natural User Interface (NUI) as there is a strong barrier between the user and computer. There are various hand tracking systems available on the market, but they are complex and expensive.

Julie Bakken Jepsen, et al. [9] 2015. Sign Languages of the World: Walter de Gruyter, Inc., Berlin, Boston and Ishara Press, Preston. Although a number of edited collections deal with either the languages of the world or the languages of particular regions or genetic families, only a few cover sign languages or even include a substantial amount of information on them. This handbook provides information on some 38 sign languages, including basic facts about each of the languages, structural aspects, history and culture of the Deaf communities, and history of research.

Marlon Oliveira. 2018.[10] Handshape Recognition using Principal Component Analysis and Convolutional Neural Networks applied to Sign Language. Doctoral Thesis, Dublin. Handshape recognition is an important problem in computer vision with significant societal impact. However, it is not an easy task, since hands are naturally deformable objects. Handshape recognition contains open problems, such as low accuracy or low speed, and despite a large number of proposed approaches, no solution has been found to solve these open problems. In this thesis, a new image dataset for Irish Sign Language (ISL) recognition is introduced.

Ananya Choudhury, et al. 2015. [11] A Review on Vision-Based Hand Gesture Recognition and Applications, Research Gate, pp.261-286. In the present scenario, vision based hand gesture recognition has become a highly emerging research area for the purpose of human computer interaction. Such recognition systems are deployed to serve as a replacement for the commonly used human-machine interactive devices such as keyboard, mouse, joystick etc. in real world situations.

Tao Liu, et al. 2016. [12] Sign language recognition with long short-term memory, IEEE International Conference on Image Processing, pp. 2871-2875. Sign Language Recognition (SLR) aims at translating the Sign Language (SL) into speech or text, so as to facilitate the communication between hearing-impaired people and the normal people.

Mais Yasen, et al. 2019. [13] A systematic review on hand gesture recognition techniques, challenges and applications. PeerJ Comput, Sci. http://doi.org/10.7717/peerj-cs.218 Background With the development of today's technology, and as humans tend to naturally use hand gestures in their communication process to clarify their intentions, hand gesture recognition is considered to be an important part of Human Computer Interaction (HCI), which gives computers the ability of capturing and interpreting hand gestures, and executing commands afterwards.

III. METHODOLOGY

The Hand Gesture Recognition paper follows the following methodology to build a hand gesture recognizer using OpenCV and Python.

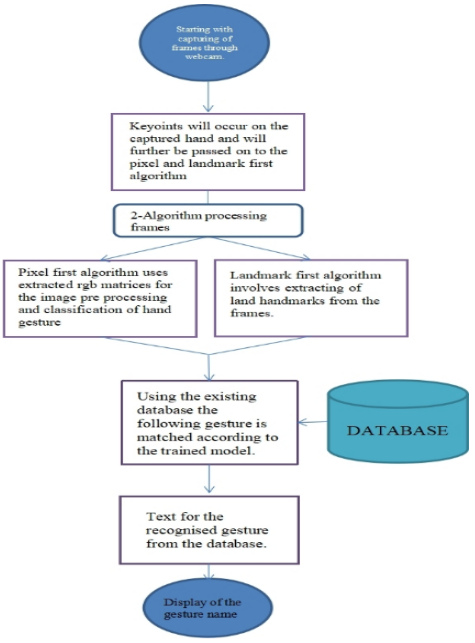

Figure 1: The sequence of processes in our methodology

1) *Preparation/Overview:* As hand gesture recognition is a vast field, the paper starts with thorough research on the topic to understand its applications, challenges, and available resources. Before starting the implementation of the hand gesture recognition paper, it is crucial to conduct thorough research on the topic. This research includes exploring the various applications of hand gesture recognition in fields such as Human-Computer Interaction, robotics, sign language translation, and music creation. The research also includes understanding the challenges associated with hand gesture recognition, such as variability in hand shapes and positions, occlusion, and environmental factors like lighting and background.

Additionally, the paper team may research the available resources and tools for building a hand gesture recognition system, such as libraries, frameworks, and datasets. This research helps in selecting the most appropriate tools and techniques for the paper and ensuring that the system performs optimally. Overall, conducting research is an essential step in any paper, including hand gesture recognition, to gain knowledge and insights that help in making informed decisions throughout the paper's development.

2) *Frameworks and libraries*: The paper uses the MediaPipe and Tensorflow frameworks for hand detection and gesture recognition, respectively, and the OpenCV library for image processing.

In this paper, MediaPipe and Tensorflow frameworks are used for detecting the hand and recognizing the gesture, respectively. MediaPipe is a cross-platform framework that provides real-time processing of multimedia data, including video, audio, and 3D data. MediaPipe's Hand Tracking pipeline is used in this paper to detect the hand and extract hand landmarks.

Tensorflow, on the other hand, is an open-source machine learning framework developed by Google. It provides a comprehensive set of tools and libraries for building and training machine learning models. In this paper, Tensorflow is used to train the neural network for gesture recognition.

In addition to these frameworks, the paper also utilizes the OpenCV library for image processing. OpenCV is an open-source computer vision library that provides a wide range of functions and algorithms for image and video processing. It is used in this paper for processing camera frames, including resizing, normalisation, and filtering.

The use of these frameworks and libraries makes it easier to develop the Hand Gesture Recognition system, as they provide pre-built components and tools that can be used to build the system. Additionally, they enable the system to run in real-time, making it suitable for interactive applications.

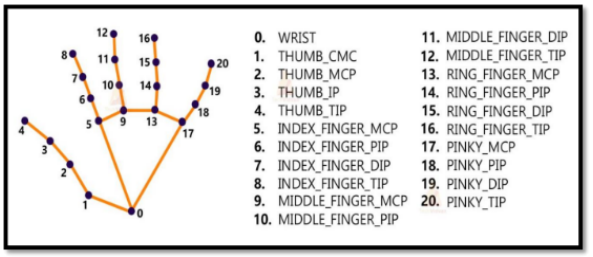

Figure 2: The Key Points

3) *Data Collection:* Data collection is an essential step in building a machine learning model, and it plays a crucial role in the accuracy and reliability of the final model. In this Hand Gesture Recognition paper, data is collected through the camera frames that are processed by OpenCV. OpenCV is a computer vision library that provides various functions and algorithms to process images and videos.The camera captures frames of hand gestures, which are then processed by OpenCV to extract features and convert them into a format that can be fed to the model. These features are then used to train the model so that it can recognize and classify the different hand gestures accurately.

The data collection process involves capturing a

large number of images of various hand gestures, ensuring that the data set is diverse and representative of the different hand gestures that the model is expected to recognize. The data set is then divided into training, validation, and testing sets, with the training set used to train the model, the validation set used to tune the hyperparameters of the model, and the testing set used to evaluate the performance of the final model.

Overall, the quality of the data collected is critical to the success of the model. The data set needs to be well-balanced, diverse, and representative of the different hand gestures that the model is expected to recognize. Additionally, the data needs to be preprocessed and formatted correctly before it can be fed to the model for training.

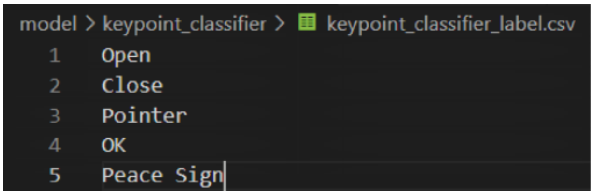

Figure 3: The Key Point Labels

4) *Training:* The training of the model is a crucial step in the hand gesture recognition paper. The two approaches used, Pixel First and Landmark First, are different in their methodology but aim to achieve the same goal: accurate classification of hand gestures.

In the Pixel First Approach, the RGB matrices of the frames are used as input to a convolutional neural network (CNN). The CNN is a type of neural network that is optimized for image recognition tasks. The network is trained to recognize patterns in the RGB matrices and classify them into different hand gestures. The training process involves feeding the network with a large dataset of labeled images, and the network adjusts its parameters to minimize the error between the predicted and actual labels.

In the Landmark First Approach, a control network is used to extract hand landmarks from the frames. The landmarks are then used as input to a simpler neural network architecture for classification. This approach requires a smaller dataset compared to the Pixel First Approach, as the network only needs to learn to recognize the hand landmarks instead of analyzing the entire image.

Both approaches have their advantages and disadvantages, and the choice of which approach to use depends on the specific requirements of the paper. However, both approaches require a significant amount of training data to achieve accurate classification results.

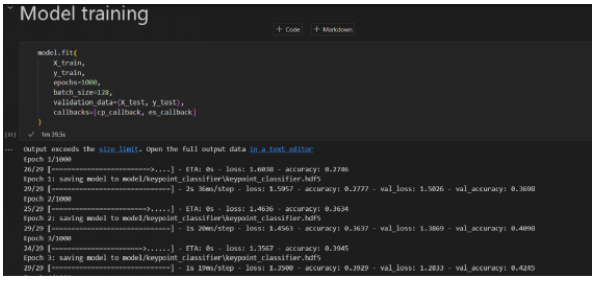

Figure 4: The Model Training

5) *Model evaluation:* The trained models are evaluated using various metrics to assess their performance. For example, if the system is designed to recognize sign language gestures, it is important to ensure that the model has high accuracy and precision in detecting and classifying different signs. In this case, false positives and false negatives can have significant consequences for the user, and therefore the model's performance should be carefully evaluated and optimized.

Model evaluation can also help to identify potential issues with the data, such as class imbalance or data bias, which can affect the performance of the model. By identifying these issues, the system can be improved to ensure that it is robust and accurate in real-world scenarios.

Overall, model evaluation is a critical step in the development of a hand gesture recognition

system, and should be performed regularly to ensure that the system is performing at its best.

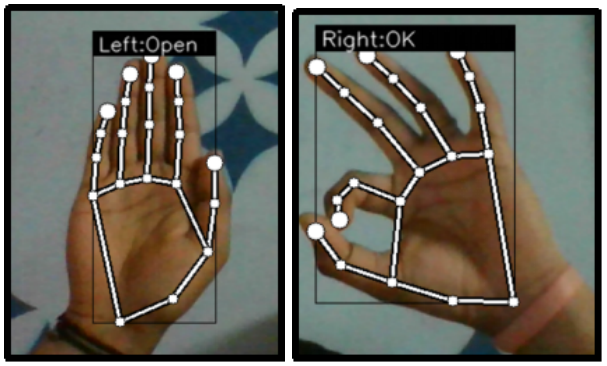

Figure 5: The Hand Gestures

6) Deployment is the final step in the development of a hand gesture recognition system. Once the model is trained and evaluated, it can be deployed for practical use. This involves integrating the model into a larger system or application, which can be used by the target users.

For example, the hand gesture recognition system can be deployed as a software application on a computer or mobile device, which can interpret hand gestures in real-time and provide corresponding outputs. This can help people with hearing and speech disabilities to communicate more effectively.

It is important to ensure that the system is robust and accurate before deployment, and that it is designed with the user's needs in mind. The deployment process also involves continuous testing and monitoring to ensure that the system is functioning as intended and meeting user expectations.

IV. EXPERIMENTAL RESULTS

The database can be created by either collecting data from real-world scenarios or by generating synthetic data. In real-world scenarios, data is collected by recording videos of people performing different hand gestures. In synthetic data generation, hand models are used to create artificial hand gestures in a controlled environment.

The quality and quantity of the database play a crucial role in the accuracy and robustness of the hand gesture recognition system. Therefore, it is important to have a well-designed database that can effectively train and test the model.

As the results of the hand gesture recognition system depend on various factors such as the quality of the dataset, preprocessing techniques used, the accuracy of the model, and the performance metrics used for evaluation, it's difficult to provide a generalised result.

However, typically, the accuracy of the hand gesture recognition system is measured in terms of its ability to correctly classify a given gesture. For example, if the system is trained on a dataset of hand gestures containing the signs for letters A, B, and C, and it correctly classifies 90% of the test data, it means that 9 out of 10 times, it is able to accurately recognize the hand gesture and classify it as the correct sign.

So, here in our paper the performance of the hand gesture recognition system is measured using metrics such as precision, recall, F1 score which give a more detailed picture of the system's performance.

In hand gesture recognition systems, accuracy is typically measured using the following equations:

True Positive (TP): The number of correct predictions of a particular gesture.

False Positive (FP): The number of incorrect predictions of a particular gesture.

False Negative (FN): The number of times a gesture is not detected.

True Negative (TN): The number of times a non-gesture is correctly identified as such.

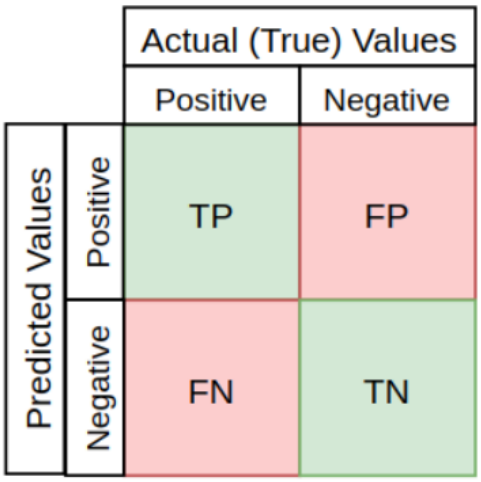

Figure 6: Accuracy measuring

Using these values, the accuracy of a hand gesture recognition system can be calculated using the following formulas:

Precision:- Precision is a measure of how many of the positive predictions made are correct (true positives). The formula for it is:

$$Precision = \frac{TP}{TP + FP}$$

Recall:- Recall is a measure of how many of the positive cases the classifier correctly predicted, over all the positive cases in the data. It is sometimes also referred to as Sensitivity. The formula for it is:

$$Recall = \frac{TP}{TP + FN}$$

F1 score:- F1-Score is a measure combining both precision and recall. It is generally described as the harmonic mean of the two. Harmonic mean is just another way to calculate an "average" of values, generally described as more suitable for ratios (such as precision and recall) than the traditional arithmetic mean. The formula used for F1-score in this case is:

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

Accuracy:- The base metric used for model evaluation is often Accuracy, describing the number of correct predictions over all predictions:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

These metrics provide a quantitative assessment of the performance of a hand gesture recognition system and are commonly used to compare different approaches or evaluate the effectiveness of improvements to an existing system.

Confusion Matrix

Gestures	Precision	Recall	F1-Score	Accuracy
Open	0.94	0.96	0.95	94%
Close	0.93	0.90	0.91	93%
Pointer	0.89	0.98	0.94	89%
Okay	0.91	0.91	0.91	91%
W	0.74	1.00	0.85	74%
Y	1.00	0.95	0.97	100%
A	0.71	0.64	0.67	71%
B	0.89	0.99	0.93	89%
C	0.99	0.84	0.91	99%

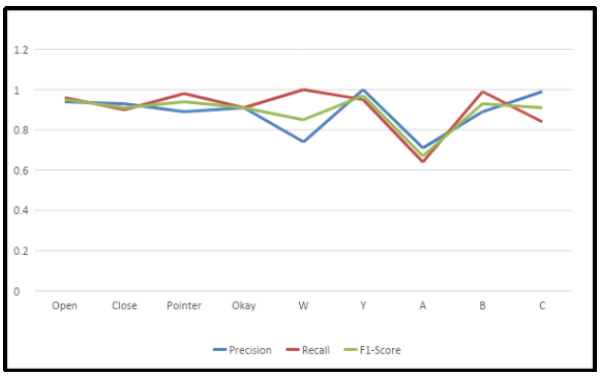

Figure 7: Confusion Matrix

OUTPUTS

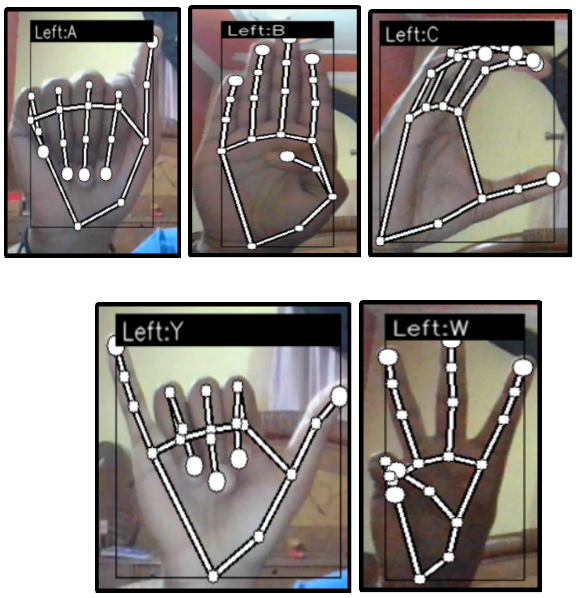

Figure 8: Different Hand Gestures with their labelled name

VI. CONCLUSION

In conclusion, in today's world, the hand gesture recognition system is rapidly emerging as a promising technology that can bridge the communication gap between the deaf or mute community and the rest of the world. The system is designed to detect and translate hand gestures into text, thereby providing an effective means of communication that is easy, accessible, and reliable. The hand gesture recognition system operates by utilizing advanced computer vision techniques, neural networks, and machine learning algorithms to accurately classify and detect hand gestures.

The effectiveness of the hand gesture recognition system largely depends on the accuracy of the models used in the system. To achieve high accuracy, proper data collection, preprocessing, training, and evaluation of the models are essential. The system can be trained on a large dataset of hand gestures, which can be further processed using common image processing techniques to enhance the quality of the images. The trained models can be evaluated using various metrics to assess their performance.

The deployment of the hand gesture recognition system can greatly benefit the deaf or mute community, as it provides them with a means of communication that is easy to use and understand. This technology can enhance their quality of life by allowing them to communicate effectively with others. As technology continues to evolve, we can expect to see more advanced and efficient hand gesture recognition systems emerge, which have the potential to make a positive impact on people's lives. In conclusion, the hand gesture recognition system is an exciting technology that has a promising future in the field of communication and accessibility.

VII. REFERENCES

[1] Kazuhito Takahashi et al. 2020 has developed a system using mediapipe, openCV and various deep learning algorithm such as Keras Neural Network and Pixels first algorithm and landmark first algorithm

[2] Rupesh Prajapati, et al. “Hand Gesture Recognition and Voice Conversion for Deaf and Dumb”

[3] Raimundo F. Pinto , et al. have proposed a sign recognition system based on CNN, ANN, Contour Extraction and polygon approximation for extraction of a feature and the image processing part.

[4] Sood, Anchal, and Anju Mishra. "AAWAAZ: A communication system for deaf and dumb." Reliability, Infocom Technologies and Optimization (Trends and Future

Directions)(ICRITO), 2016 5th International Conference on. IEEE,2016.

[5] Karray F., et al. 2008 Human- Computer Interaction: Overview on State of the Art Int. J. Smart Sens. Intell. Syst. 1 137-159.

[6] Murthy G. et al. 2009 A review of vision based hand gestures recognition Int. J. Inf. Technol. Knowl. Manag. 2 405-410.

[7] Garg P., et al. 2009 Vision based hand gesture recognition World Acad. Sci. Eng. Technol. 972-977.

[8] Yeo H.-S., et al. 2013 Hand tracking and gesture recognition system for human- computer interaction using low-cost hardware Multimed. Tools Appl.

[9] Julie Bakken Jepsen, et al. 2015. Sign Languages of the World: Walter de Gruyter, Inc., Berlin, Boston and Ishara Press, Preston.

[10] Marlon Oliveira. 2018. Handshape Recognition using Principal Component Analysis and Convolutional Neural Networks applied to Sign Language.

[11] Ananya Choudhury, et al. 2015. A Review on Vision-Based Hand Gesture Recognition and Applications, Research Gate, pp.261-286.

[12] Tao Liu, et al. 2016. Sign language recognition with long short-term memory, IEEE International Conference on Image Processing, pp. 2871-2875.

[13] Mais Yasen, et al. 2019. A systematic review on hand gesture recognition techniques, challenges and applications. PeerJ Comput, Sci. http://doi.org/10.7717/peerj-cs.218