IntellectSoft: Intelligence which works for Physically Disabled People

Prof. Swapnil S. Chaudhari
Computer Department
MMIT, Lohegaon Pune

Ms. Ruchika Bhoite
Computer Department
MMIT, Lohegaon Pune

Ms. Prachi Shinde
Computer Department
MMIT, Lohegaon Pune

Ms. Rakhi Manaware
Computer Department
MMIT, Lohegaon Pune

Ms. Vaishnavi Mahajan
Computer Department
MMIT, Lohegaon Pune

Abstract— Hand gesture is one of the methods used in sign language for non-verbal communication. It is most commonly used by deaf & dumb people who have hearing or speech problems to communicate among themselves or with normal people. Various sign language systems had been developed by many makers around the world but they are neither flexible nor cost-effective for the end users. Hence, it is a software which presents a system prototype that is able to automatically recognize sign language to help deaf and dumb people to communicate more effectively with each other or normal people. Dumb people are usually deprived of normal communication with other people in the society, also normal people find it difficult to understand and communicate with them. These people have to rely on an interpreter or on some sort of visual communication. An interpreter won't be always available and visual communication is mostly difficult to understand. Sign Language is the primary means of communication in the deaf and dumb community. As a normal person is unaware of the grammar or meaning of various gestures that are part of a sign language, it is primarily limited to their families and/or deaf and dumb community.

Keywords: HCI (Human-Computer Interaction), HSV (Hue Saturation Value), Gesture Recognition, Computer Graphics, Layout's, Rendering, CNN (Convolutional Neural Network) & KKN (k-nearest neighbors' algorithm)

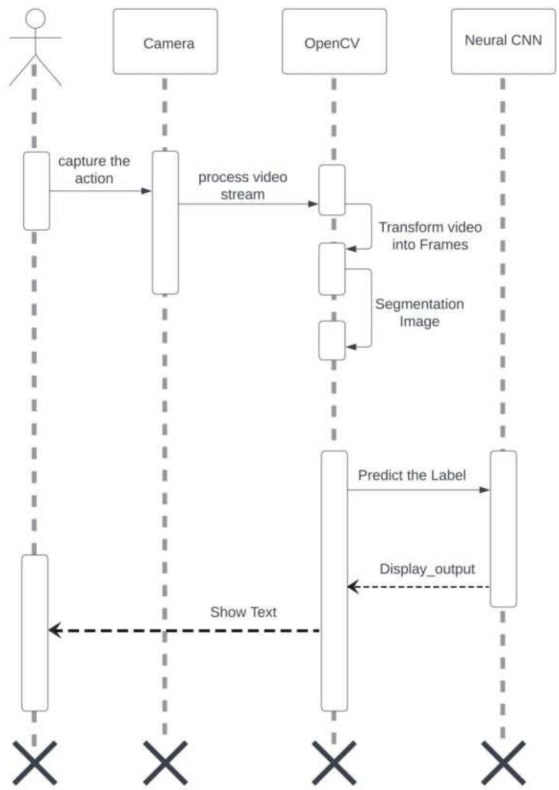

Fig 1: Sequence diagram

1. INTRODUCTION

The goal of this project was to build a neural network able to classify which letter of the American Sign Language (ASL) alphabet is being signed, given an image of a signing hand. This project is a first step towards building a possible sign language translator, which can take communications in sign language and translate them into written and oral language. Such a translator would greatly lower the barrier for many deaf and mute individuals to be able to better communicate with others in day-to-day interactions. This goal is further motivated by the isolation that is felt within the deaf community. Loneliness and depression exist in higher rates among the deaf population, especially when they are immersed in a hearing world [1]. Large barriers that profoundly affect life quality stem from the communication disconnect between the deaf and the hearing. Some examples are information deprivation, limitation of social connections, and difficulty integrating in society [2].

Most research implementations for this task have used depth maps generated by depth camera and high-resolution images. The objective of this project was to see if neural networks are able to classify signed ASL letters using simple images of hands taken with a personal device such as a laptop webcam. This is in alignment with the motivation as this would make a future implementation of a real time ASL-to-oral/written language translator practical in an everyday situation

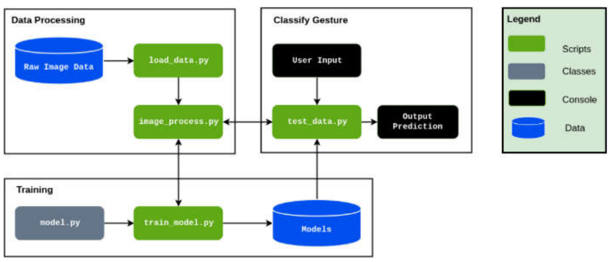

Fig 2: Block Diagram of Software

As shown in Figure 1, the project will be structured into 3 distinct functional blocks, Data Processing, Training, Classify Gesture. The block diagram is simplified in detail to abstract some of the minutiae:

• Data Processing: The load data.py script contains functions to load the Raw Image Data and save the image data as numpy arrays into file storage. The process data.py script will load the image data from data.npy and preprocess the image by resizing/rescaling the image, and applying filters and ZCA whitening to enhance features. During training the processed image data was split into training, validation, and testing data and written to storage. Training also involves a load dataset.py script that loads the relevant data split into a Dataset class. For use of the trained model in classifying gestures, an individual image is loaded and processed from the filesystem.

• Training: The training loop for the model is contained in train model.py. The model is trained with hyperparameters obtained from a config file that lists the learning rate, batch size, image filtering, and number of epochs. The configuration used to train the model is saved along with the model architecture for future evaluation and tweaking for improved results. Within the training loop, the training and validation datasets are loaded as Dataloaders and the model is trained using Adam Optimizer with Cross Entropy Loss. The model is evaluated every epoch on the validation set and the model with best validation accuracy is saved to storage for further evaluation and use. Upon finishing training, the training and validation error and loss is saved to the disk, along with a plot of error and loss over training.

• Classify Gesture: After a model has been trained, it can be used to classify a new ASL gesture that is available as a file on the filesystem. The user inputs the file path of the gesture image and the test data.py script will pass the file path to process data.py to load and preprocess the file the same way as the model has been trained.

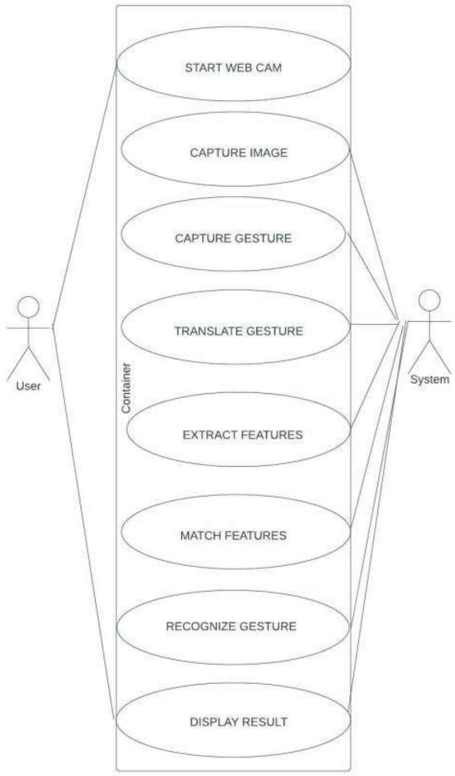

Fig 3: Use case diagram

1. LITERATURE REVIEW

[1] In this paper, American Sign Language Words Recognition Using Spatio-Temporal Prosodic and Angle Features: A Sequential Learning Approach, in this article, we adopted an approach to recognize highly correlated American sign language words. We optimize the accuracy of recorded 3D video skeletal hand joints information, using a WLR algorithm and filter.

[2] In this paper, End-to-End Dynamic Gesture Recognition Using MmWave Radar, Millimeter-wave (mmWave) radar sensors are a promising modality for gesture recognition as they can overcome several limitations of optic sensors typically used for gesture recognition. These limitations include cost, battery consumption, and privacy concerns.

[3] In this paper, An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion, this approach preserves the spatial and temporal information of the sign by fusing the sign's key postures in the forward and backward directions to generate an accumulative video motion frame.

[4]. In this paper, Facial Expression Recognition Using Hybrid Features of Pixel and Geometry This article presents a SACNN-ALSTMs to extract the hybrid feature for facial expression recognition. The SACNN-ALSTMs can learn the relative geometric position dependencies of facial landmark points and extract more discriminative facial features for FER.

2. PROPOSED SYSTEM

The recognition of sign language gestures from real time video and successfully classifying it into either one from a list of categories have been a popular and challenging field of research. Many researchers have been working on this field for a long time, so we have also thought of contributing to this field as by working on it in our final year major project. *Liang et al. [6]* have also put their research on this concept which has guided us throughout the implementation. The process of recognizing a sign language gesture and classifying it is the one line definition of the task performed by this proposed system. Along with this, a text to ASL finger spelling feature is also available that makes the two-way communication from sign to text and text to sign possible.

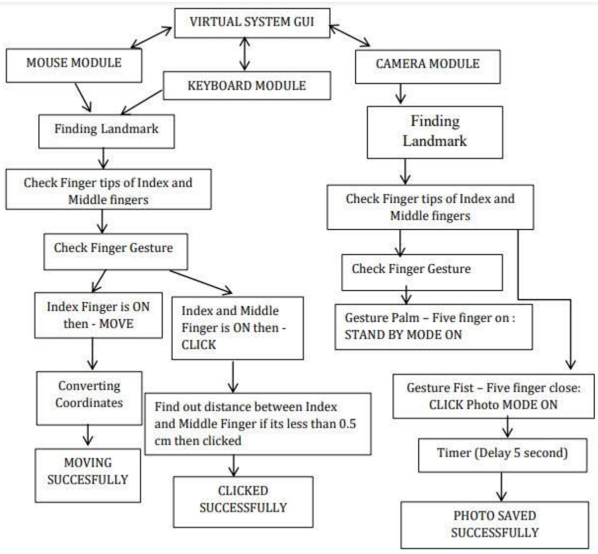

Fig 4: System architecture

Above Fig.2;shows the Architecture of IntellectSoft system that how IntellectSoft works from inside the system. In these several process take place as Input Image, Image Resize, Segmentation, Get Radius, Get Fingertips, Controlling are done inside the IntellectSoft.

3. ADDITIONAL FEATURES

A. *SIGN LANGUAGE:*

Sign Language consists of fingerspelling, which spells out words character by character, and word level association which involves hand gestures that convey the word meaning. Fingerspelling is a vital tool in sign language, as it enables the communication of names, addresses and other words that do not carry a meaning in word level association. In spite of this, fingerspelling is not widely used as it is challenging to understand and difficult to use. Moreover, there is no universal sign language and very few people know it, which makes it an inadequate alternative for communication.

A system for sign language recognition that classifies finger spelling can solve this problem. Various machine learning algorithms are used and their accuracies are recorded and compared in this report.

American Sign Language (ASL) dataset created by B. Kang et al is used. It is a collection of 31,000 images, 1000 images for each of the 31 classes. These gestures are recorded for a total of five subjects. The gestures include numerals 1- 9 and alphabets A-Z except ‘J’ and ‘Z’, because these require movements of hand and thus cannot be captured in the form of an image. Some of the gestures are very similar, (0/o) , (V/2) and (W/6). These are classified by context or meaning.

Fig 5: Basic Sign Language

B. *INPUT IMAGE:*

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

1. Importing the image via image acquisition tools;
2. Analysing and manipulating the image;
3. Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

4. CONCLUSION

The main objective of the AI IntellectSoft is to control the mouse cursor, keyboard and camera functions by using the hand gestures instead of using a physical hardware. The proposed system can be achieved by using a webcam or a built-in camera which detects the hand gestures and hand tip and processes these frames to perform the particular mouse functions.

From the results of the model, we can come to a conclusion that the proposed AI IntellectSoft has performed very well and has a greater accuracy compared to the existing models and also the model overcomes most of the limitations of the existing systems. Since the proposed model has greater accuracy, the AI IntellectSoft can be used for real-world applications, since the proposed system can be used virtually using hand gestures without using the traditional physical touch.The model has some limitations such as small decrease in accuracy in click keyboard function and some difficulties in clicking and dragging to select the text. Hence, we will work next to overcome these limitations by improving the finger-tip detection algorithm to produce more accurate results.

5. ACKNOWLEDGMENT

In the accomplishment of this project successfully, many people have extended a helping hand, we would like to show our deep appreciation for them and we are utilizing this time to thank all the people who have been concerned with the project. I would like to thank our Principal, Dr. R. V. Bortake, and our HOD, Prof. Subhash G. Rathod, for providing us with the golden opportunity to work on this project. Their suggestions and instructions have served as the major contribution towards the completion of this project. We would also like to thank our guide, Prof. Swapnil S. Chaudhari, for his valuable support and guidance throughout the completion of our project. We would also like to thank our classmates who have helped us with their valuable suggestions and support which has been very helpful in the completion of this project.

6. REFERENCES

[1] (Apr. 30, 2022). Hearing Loss Statistics. [Online]. Available:798 https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing- 799 loss 800

[2] H. Luqman and S. A. Mahmoud, ‘‘Automatic translation of Arabic text- 801 to-Arabic sign language,’’ Universal Access Inf. Soc., vol. 18, no. 4, 802 pp. 939–951, 2018. 803

[3] A. A. I. Sidig, H. Luqman, and S. A. Mahmoud, ‘‘Transform-based Ara- 804 bic sign language recognition,’’ Proc. Comput. Sci., vol. 117, pp. 2–9, 805 Nov. 2017. 806

[4] A. A. I. Sidig, H. Luqman, and S. A. Mahmoud, ‘‘Arabic sign language 807 recognition using optical flow-based features and HMM,’’ in Proc. Int. 808 Conf. Reliable Inf. Commun. Technol., 2017, pp. 297–305. 809

[5] LAS: Second Part of the Unified Arabic Sign Dictionary, League Arab 810 States Arab League Educ., Cultural Sci. Org., Tunisia, 2006. 811

[6] LAS: First Part of the Unified Arabic Sign Dictionary, League Arab States 812 & Arab League Educ., Cultural Sci. Org., Tunisia, 2000. 813

[7] G. M. B. Makhashen, H. A. Luqman, and E.-S.-M. El-Alfy, ‘‘Using Gabor 814 filter bank with downsampling and SVM for visual sign language alphabet 815 recognition,’’ in Proc. 2nd Smart Cities Symp. (SCS), 2019, pp. 1–6. 816

[8] A. V. Nair and V. Bindu, ‘‘A review on Indian sign language recognition,’’ 817 Int. J. Comput. Appl., vol. 73, no. 22, pp. 33–38, 2013. 818

[9] M. J. Cheok, Z. Omar, and M. H. Jaward, ‘‘A review of hand gesture 819 and sign language recognition techniques,’’ Int. J. Mach. Learn. Cybern., 820 vol. 10, pp. 131–153, Aug. 2019. 821

[10] Kong Saoane Thach, “User’s perception on mental health applications”, School of Engineering and Technology, Tra Vinh University, Tra Vinh, Viet Nam