



DEEP LEARNING METHODS TO DETECT HAND GESTURES FOR DEAF AND DUMB

¹K Vasudha, Assistant Professor, Guru Nanak Institutions Technical Campus

²N Sudha Laxmaiah, Assistant Professor, Bhoj Reddy Engineering College for Women

³A Hima Bindu, Assistant Professor, Bhoj Reddy Engineering College for Women

Abstract: In Recent years human-machine interaction has become the most interested field in computer science & engineering. Artificial intelligence and Deep learning have advanced in leaps and boundaries, thanks to the rapid growth of hardware and software technologies. Speech and Gesture are the two key functionalities for Human Machine Interaction. The main focus here is on extracting features from images and identify the gestures. In this paper we explored Convolutional Neural Networks (CNN) for this purpose. CNN set up deep Learning methods for completing the task. In this paper, we proposed a model to detect hand gestures from the trained dataset of given images. From an each image, fingers and palms are separated so that the gestures can be interpreted easily. we used a dataset of nearly 2500 images so that the margin of error could be reduced. images are captured using a simple camera (preferably laptop web-cam for collecting images) and the size of the dataset may become very large because images need to be captured from different angles, to be able to identify a particular gesture. The outcome of results are compared with the preexisting results of KNN and HMM models. In this paper we did only exploring CNN in depth. because they could make hand gesture recognition a cost-effective process. later this can be facilitate future work for remote learning for deaf and dumb community across the globe.

Keywords — *Convolutional Neural Network, Hand Gestures, Human Machine Interaction, Image Preprocessing, Computer Vision.*

I. INTRODUCTION

Hand Gesture recognition is a process that wants to decipher human gestures with the help of mathematical processes. HMI is not limited to just keyboard and mouse use. It has gained vast functionalities due to technical advancements in the previous decade. Hand gestures are classified into two categories: Static and dynamic gestures. The authors have compared various static hand recognition methods like Hidden Markov Method, Convolutional Neural Networks, and K-Nearest Neighbors method. The motivation for choosing this area of research was the fact that the deaf and dumb community all over the world, especially in India do not have a universal way of communicating and they are not accorded the respect they deserve. The authors have felt the need to create a methodological way so that these people

are enabled to talk to people easily and they could also receive quality education with the help of chat bots that could be designed to create an interactive platform for these people. In this paper, the authors have thoroughly compared various pre existing technologies that have been employed for creating a hand gesture recognition platform. An elaborative literature review has also been done to learn about these methods in detail.

➤ DEEP LEARNING

Deep Learning is a subset of machine learning that mainly a neural network which has 3 or more layers. It is also a subgroup of Machine Learning. These neural networks try to copy the functioning of the human brain by training themselves (manually or automatically) using enormous amounts of dataset. Multiple layers are recommended so that the machine can make more very accurate predictions. Even though a single neural network is usable, it is not accurate enough. Deep Learning has gained huge traction lately due to the fact that it eliminates human intervention making the process fast and precise.

➤ CONVOLUTIONAL NEURAL NETWORKS

Majority of computations takes place in convolutional layer, hence it is the most important component of a CNN. Its components are input data, filter, and features map. Multiple convolutional layers form the architecture of convolutional neural networks. This architecture is very similar to the pattern of neurons in the human brain and therefore the name "Neural Network". The algorithm of CNN works as follows : An

image is taken as an input, it is accorded weights and biases relative to its quality, and then it is used as a dataset according to the requirements of a particular project.

➤ FEATURES FOR GESTURE RECOGNITION

Selecting features is crucial to gesture recognition, since hand gestures are very rich in shape variation, motion and textures. For static hand posture recognition, although it is possible to recognize hand posture by extracting some geometric features such as fingertips, finger directions and hand contours, such features are not always available and reliable due to self occlusion and lighting conditions. There are also many other non-geometric features such as color, Silhouette and textures, however, they are inadequate in recognition. Since it is not easy to specify features explicitly, the whole image or transformed image is taken as the input and features are selected implicitly and automatically by the recognizer. Hand features can be derived using the following two approaches:

(a) **Model based Approaches** (Kinematic Model): Model based approaches attempt to infer the pose of the palm and the joint angles [4, 5]. Such an approach would be ideal for realistic interactions in virtual environments. Generally, the approach consists of searching for the kinematic parameters that brings the 2D projection of a 3D model of hand into correspondence with an edge-based image of a hand [6]. A common problem with the model-based approaches is the problem of the feature extraction (i.e. edges). The human hand itself is rather texture less and does not provide

many reliable edges internally. The edges that are extracted are usually extracted from the occluding boundaries. In order to facilitate extraction and unambiguous correspondence of edges with model edges the approaches require homogeneous backgrounds and high contrast backgrounds.

(b) **Low Level Features based Approaches:** In many gesture applications though all that is required is a mapping between input video and gesture. Therefore, many have argued that the full reconstruction of the hand is not essential for gesture recognition. Instead many approaches have utilized the extraction of low-level image measurements that are fairly robust to noise and can be extracted quickly. Low-level features that have been proposed in the literature include: the centroid of the hand region [7], principle axes defining an elliptical bounding region of the hand, and the optical flow/affine flow of the hand region in a scene.

➤ CLASSIFICATION OF HAND GESTURES

Hand Gestures can be classified using the following two approaches.

(a) **Rule based Approaches:** Rule-based approaches consist of a set of manually encoded rules between feature inputs. Given an input gesture a set of features are extracted and compared to the encoded rules, the rule that matches the input is outputted as the gesture. As an example, in [9] predicates related to low-level features of the motion of the hands are defined for each of the actions under consideration. When a predicate of a gesture is satisfied over a fixed number of consecutive frames the gesture is

returned. A major problem with rule-based approaches is that they rely on the ability of a human to encode rules.

(b) **Machine Learning based Approaches:** A popular machine learning approach is to treat a gesture as the out-put of a stochastic process. Of this class of approach Hidden Markov Models (HMMs) [10,11] by far have received the most attention in the literature for classifying gestures.

II. RELATED WORK

In this paper, the authors have focused on the working of ConvNet and compared our results with the existing results for KNN and HMM models. In their study,

- Thad Starner and Alex Pentland[1] looked at two real-time Hidden Markov Model systems. They utilized it to recognise ASL at the sentence level. Sentences with the correct grammar showed higher accuracy than the ones without it. As this work was done prior to 1997, they used a simple camera to capture the images.
- In their publication, Robyn Owens, Gareth Lee, and Eun-Jung Holden (2005) [2] refined the HMM by tracking several target objects and improved gesture recognition accuracy. They used Australian Sign Language to execute their work (Auslan). They were able to attain a sentence level accuracy of 97 percent and a word level accuracy of 99 percent.
- K. Imagawa, Shan Lu, S. Igi (1998) [3] described a real-time system that is capable of tracking hands even when they are overlapping

the face. This system was tested by native signers. It employed skin color differentiation to distinguish between face and hand from images.

- Razieh Rastgoo, Kourosh Kiani & Sergio Escalera (2020) [4] further improved upon the CNN based model and used 3D hand key points and 2D input frames. They also created an enormous amount of dataset to aid their research work. This model proved to be more efficient than any of the previous models thereby paving a way for logical, efficient and cost-effective way for use of CNN for gesture recognition.

III. TECHNOLOGIES

➤ PYTHON

At the moment, Python is the widely used programming language in the world that is used to build software and websites, automate tasks, and it is the most prominent. Some of the most important traits that make Python user friendly are : active and a large community, being open source and most important having all the requisite functions as predefined libraries. Python libraries used are: Tkinter GUI 8.0— Python interface to TCL/Tk, pypiwin32, pytsx3, tensor flow, keras, scipy, opencv, qimage2ndarray, pillow

➤ OPENCV

Open source Computer Vision library is the full form of OpenCV. Well established companies like Meta, Google, Microsoft, Toyota and Intel make extensive use of OpenCV. It was created to provide a common infrastructure for CV and speed up use of Machine Learning[9] across the globe. OpenCV

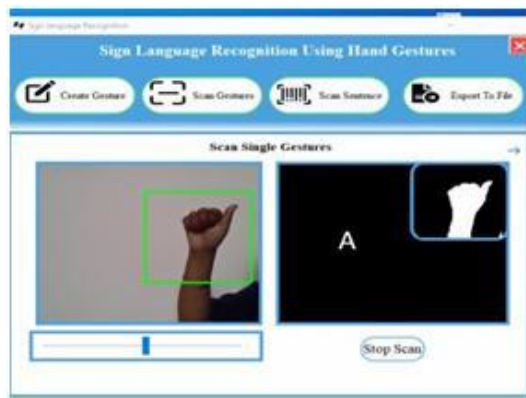
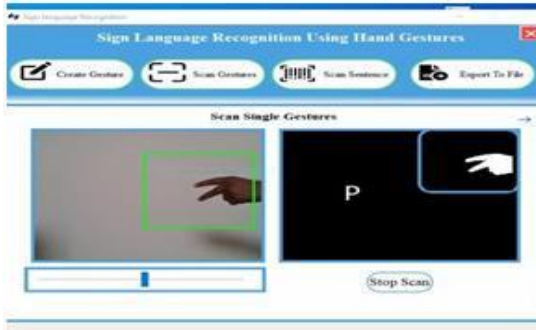
supports Linux, Windows, and all other operating systems and has C++, Python, and MATLAB interfaces. Its libraries contain over 2000 algorithms for accelerating operability. It has become the most used source for performing Computer Vision tasks and is extensively used for research work. It has a community of active users that helps drive its usage more than any other software for Computer Vision.

IV. METHODOLOGY

➤ ALGORITHM

Generally, poses and gestures are one the basic means of communication between humans and systems as they are able to transfer some kind of meaning. The examination region of posture furthermore expects to perceive such articulations, which regularly include some stance or potential movement of the hands, arms, head, or even skeletal joints of the entire body. In gesture recognition, there are certain image processing features, which are trailed by neural systems to help order the distinguished motion. Image Preprocessing involves various processes like preprocessing can be done using two modes: Binary and Skin Mask mode, then noise removal is done on these images using Gaussian Blur Method[10] and Dilation Methods. To achieve better results, the images are subjected to additional image segmentation, enhancement, and morphological filtering.

V. OUTPUT SCREENS



VI. LIMITATIONS AND CONCLUSION

In this paper, for the American Sign Language[10] classification method, the authors followed a deep learning strategy. The major drawback of this model is that the authors would need different types of datasets to process various types of gestures.

This might prove fatal as they could encounter lack of space for storing these datasets. Also, this model will be very difficult to understand thereby making future developments tougher and requiring a good amount of prior knowledge. The significance of gesture recognition can be seen in the development of effective human-machine interaction. The system is set up depending on the photographs that have been recorded. The authors ultimately aim to create a holistic and comprehensive representation learning system for which they have designed a set of traits that can be recognized from simple gesture photographs that will help to speed up translation.

VII. FUTURE SCOPE

While the researchers cannot reach a goal of 100% accuracy in gesture recognition, they can end up creating a system that can get extremely near to that goal if you have adequate time and data. As with any effort of this nature, there is potential for improvement. The dataset contains further room for development. As the size of the dataset grows larger, the algorithms' precision will improve. Hence, more data will surely make the model more accurate and in future, and also provide a way for recognizing Indian Sign Language(ISL).

REFERENCES

- [1] Juan P. Wachs, Helman Stern, Yael Edan, “Automated Setup of a Hand-Gesture Recognition System” IEEE Transactions on Systems, Man, and Cybernetics-Part A:Systems and Humans, 35(6) (2005) 932-944.
- [2] Stenger, B., Mendonca, P. & Cipolla, R. “Model-Based 3D Tracking of an Articulated Hand”. In IEEE Conference on Computer Vision and Pattern Recognition, (2001) 310– 315.
- [3] Wu, Y., Lin, J. and Huang, T. “Capturing Natural Hand Articulation”.In IEEE International Conference on Computer Vision II (2001) 426–432.
- [4] New, J. R., Hasanbelliu, E. and Aguilar, M. “Facilitating User Interaction with Complex Systems via Hand Gesture Recognition.” In Proc. of Southeastern ACM Conf., Savannah, (2003).
- [5] M. H. Yang, N. Ahuja, and M. Tabb, “Extraction of 2- D Motion Trajectories and its Application to Hand Gesture Recognition,” in PAMI., 29(8) (2002) 1062–1074.
- [6] Cutler, R. & Turk, M. “View-Based Interpretation of Real-Time Optical Flow for Gesture Recognition”. In IEEE Int. Conf. on Automatic Face and Gesture Recognition, (1998) 416–421.
- [7] Mu-Chun Su, “A Fuzzy Rule-Based Approach to Spatio-Temporal Hand Gesture Recognition”, IEEE Trans. On Systems, Man and Cybernetics, Appl. and Review, 30 (2000).
- [8] Starner, T., Weaver, J. & Pentland, A. “Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video”. IEEE Trans. on PAMI, (1998) 1371–1375.
- [9] Wilson. A. & Bobick. A. “Parametric Hidden Markov Models for Gesture Recognition”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(9) (1999) 884–900.
- [10] Lee, H. & Kim, J. “An HMM-Based Threshold Model Approach for Gesture Recognition”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10) (1999) 961–973.
- [11] Starner,T.,Weaver,J.&Pentland,A.“Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video”. PAMI, 20(12) (1998) 1371–1375