IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

An Approach to Convert Sign to Text for Impaired People

B J M Ravi Kumar[1], B Pavan Kumar[2], B Nandini[3], A Devanand[4], B Ramakrishna[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Raghu Institute of Technology, Visakhapatnam, AP, India

Abstract: *Human Computer Interface is the study of how humans and computers interact. Hand gestures are a great way to communicate with people when they don't understand exactly what we are saying. Understanding hand gestures is essential to make sure the listener understands what we are saying. The main idea of our project is to try different approaches to hand gesture recognition. This proposed work first with radar data and then with camera sensor to achieve hand gesture recognition. First, we tried to build hand gesture recognition using radar data, and since most people don't know sign language and few interpreters, we developed an approach to real-time approach for American Sign Language based on neural networks finger spelling followed by another model with Media Pipe. We propose a complex neural network method to detect hand gestures of human behaviour from camera recorded images. The hand gesture first goes through the filter and after applying the filter the gesture goes through a classifier that predicts which type of hand gesture it is. In an existing system radar unable to detect static gestures in our approach, a deep learning-based image captioning algorithm captures both static and dynamic gestures through Media Pipe.*

Keywords: Deep Learning Techniques, CNN, Radar, Camera, Media Pipe, Sign Language

I. INTRODUCTION

Over the previous few decades, mobile and computing technologies have advanced dramatically. In addition, the way people engage with machines has improved. Human– computer interactions (HCI) approaches range from simple keyboard inputs to advanced vision-based gesture detection systems. Hand gesture recognition is one of the most exciting and important HCI technologies. Hand gesture recognition is an intriguing research subject since it may be used to improve communication in a number of applications such as mobile phones, smart TVs, robot controllers, medical equipment, access control systems, and smart automobiles. Few Applications are: On Air Interaction with Computer, Medical Field, Gesture Based Gaming Control, Autonomous Navigation System, Controlling Smart Gadgets, Communication.

Over the last decade, there has been an increasing interest in developing HCI based on Hand Gesture Recognition employing radar and other RF sensors (HGR). RADAR can penetrate clouds, fogs, mist, and insulators. It is capable of determining the precise position of an item. It can compute the velocity of a target. It can calculate the distance between two objects. Radar has numerous benefits over other sensors, including the ability to discern between fixed and moving objects. As a result, we will be utilizing radar acquired data for this project. People with speech impairments must rely on sign language since they are unable to communicate using their hearing or voice. Everyone who is speech handicapped uses sign language, but they have a difficult time interacting with non-signers (those who aren't fluent in sign language). For persons who are deaf or hard of hearing, a sign language interpreter is an essential. Their informal and formal communication is hampered as a result of this Recent advancements in deep learning have resulted in significant improvement in the fields of gesture recognition and motion recognition. In real time, the proposed technique attempts to convert hand gestures into comparable English text. This approach captures hand gestures on video and turns them into text that a non-signer can comprehend. Previously, similar research was conducted, with the bulk of them focused solely on sign translation of English alphabets or numerals.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

1.1 Project Deliverables

Hand gestures will be classified using the CNN algorithm. This technology will bridge the communication gap between signers and non-signers. This will make communication simpler 2 for people who have difficulty speaking. American Sign Language is the dominant sign language Since the only disability that D&M people have is the ability to communicate and they cannot use spoken language; their only way of communicating is through sign language. Communication is the process of exchanging thoughts and messages in different ways such as words, signals, behaviour and images. Deaf and hard of hearing persons (D&M) utilize their hands to create various gestures to communicate their ideas to others. Gestures are nonverbal communications that are communicated and are understood visually.

1.2 Scope of the Project

The main aim of the study is to recognize different hand gestures using different approaches The proposed method attempts to translate hand movements to the English text equivalent unreal time. This method uses video to record hand movements and convert them into text that can understood by non-signers. Similar Studies have been done before, focusing mainly on translating the signs of algorithm will be used to classify the hand movements. The communication gap between signatories and non-signer's will be bridged through this technology. This will facilitate communication for the visually impaired..

II. RELATED WORK

In R. R, Subramanian [1] approach the frequency-modulated continuous wave (FMCW) radar that can detect frequency shifts between transmitted and received electromagnetic waves according to the Doppler effect. The detection system is mounted on a radar array consisting of three continuous wave radars operating at 24.125GHz. A decision tree algorithm is constructed and evaluated as a classifier for the experiment. As a result, gesture movements are recorded according to the Doppler effect of the frequency spectrum of the radar signal. This system has higher in-plane motion detection performance than flexion and extension, and has achieved a high detection accuracy of 92% or more.

In M. S. Murali [2] Ultrasonic Frequency Modulated Continuous Wave (FMCW) and ConvLSTM models. One transmitter and three receivers spatially installed in different directions. The FMCW signal transmitted by the transmitter is reflected by hand and then detected by the receiver. Next, we obtained a range Doppler map (RDM) of the received signal by processing a 2D fast Fourier transform. High resolution at a distance of 0.005 m and speed of 0.03 m / s for hand gestures and 85.7% accuracy are achieved with the small size of 50 training samples of finger gestures.

In R. Raja Subramanian, & Karthick Seshadri [3] we used the DCNN and VGG16 algorithms. ASL hand gesture movements are captured as microwave Doppler signals using a microwave X-band Doppler radar transceiver. These hand gestures are analysed using MATLAB. The DCNN algorithm is used to train ASL gestures represented in spectrograms. The average validation accuracy of the DCNN and VGG16 algorithms was 87.5% and 95%, respectively.

In R. Raja Subramanian, & V. Vasudevan [4], we used FMCW radar and Deep Convolutional Neural Network (DCNN). The FMCW radar operated in the 24GHz ISM frequency band, the effective isotropic radiation power level was 0 dBm, and the FMCW radar received only one channel. Gesture recognition is performed using a deep convolutional neural network trained and tested in the Micro Doppler spectrogram. After training and validation, both methods yielded 99% classification accuracy in the test set.

In H. Mohan, A. Mounika Jenny [5], we worked on FMCW radar, signal processing, and detection by a convolutional neural network (CNN). A sequence of object distance, velocity, and azimuth information was merged into a single input and sent to a convolutional neural network for learning spatial and temporal patterns. VGG10 converged to the 10th epoch with a verification accuracy of 92%. ResNet20 is superior to VGG10 with 98% verification accuracy. CNN + LSTM has the lowest accuracy in LEFT / RIGHT due to the lack of Angle Of Arrival information, and the model achieves an average accuracy of 98% in the test set.

In Raja Subramanian R, & Vasudevan V [6]. A region-based deep complex neural network (RDCNN) is proposed to detect and classify gestures measured by a frequency modulated continuous wave radar system. In addition to the signature µD, we combine the phase difference information of the signal received from the L-shaped antenna array. The input of the proposed array contains three channels, i.e., one spectrogram and two out of phase channels. Results

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

achieved 95% (96%) average PPV (APR) for nine gestures. Dop Net dataset the hand gesture radar dataset is used for gesture classification. The micro doppler signature is used as model input.

In D.Shrestha [7]. Separate convolutional neural networks are used here. In order to accumulate without overloading, the paper uses a dissectible convolutional neural network model that performs deep convolution followed by point convolution. Get 94.56 accuracy on Dop Net data. Also, the computational time is minimized by using separable convolutions. The Micro Doppler image dataset contains 15 types of sign language actions captured by radar echo as measured by the MDHandNet model.

In M. Lakshmi [8] The proposed MDHandNet model for hand gesture or sign language recognition. The accuracy obtained was 97.1%. Compared with other methods, the proposed model has good performance, fewer parameters and lower computational complexity. Radar system, Soli dataset, Dop Net dataset using Spiking neural network have been deployed.

In B.Deepak [9]. The signal to collision conversion scheme is used to encode the Doppler radar map into spike trains fed to spike neural networks. The SNN's reader signal is fed into different classifiers. The dataset 20BNjester (Open Source - Video Dataset) is used in combination between 3DCNN and LSTM (deep learning) networks.

III. EXISTING SYSTEM

- The existing system for the detection of hand gestures, there are several types of hardware and sensors available
- The system is difficult to operate due to the utilization of several sensors and hardware
- In addition, systems have relied on stereo cameras, which are more expensive and consume more resources. The current software for hand recognition is extremely sluggish and does not generate accurate results.
- This is a significant flaw in the existing system. Many consumers are having issues with the software they are utilizing.

3.1 Challenges

Latency Image processing can be significantly slow creating unacceptable latency for videogames and other similar applications.

Robustness Due to issues such as inadequate backdrop light, strong background noise, and so on, many gesture recognition systems cannot recognize motions precisely or optimally.

Lack of Gesture Language Different users make gestures differently, causing difficulty in identifying motions. Performance The image processing needed in gesture detection is highly resource expensive, making it difficult for apps to run on resource restricted device.

IV. PROPOSEDAPPROACH

For deaf and dumb persons, we suggest converting hand gestures into text in this project. The main goal of our project is to recognize hand gestures, detect gestures, and display 8 the results as text. In front of camera, the end user must make hand motions. Our application will identify the motions as they are made by the user and will convert the into text in real time. The video obtained from the camera unit will be displayed. Our project serves as a deaf and dumb translator. It solves a number of issues including the necessity for a human translation. Our program will allow deaf and dumb individuals to express themselves. We'll use the camera to identify the hand motions. To use a camera to detect these motions, we must first isolate the hand region, deleting any undesired sections from the video sequence collected by the camera. We count the fingers visible to the camera after segmenting the hand region to direct a program based on the finger count. As a result, the entire problem can be handled in five easy steps:

- We must first locate and segment the hand region in the video system. Then from the segmented hand region in the video sequence, count the number of fingers and the size of the palm.
- The segmented section will then be matched to the available dataset. Then for quicker data selection choose the most accurate data from the dataset and apply a weight for the next comparison.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

- Finally, we'll translate the data we've gathered into text and display it. In front of the camera, the end user must make hand motions.
- Our application will identify the motions as they are made by the user and will convert them text in real time.
- Our project serves as a deaf and dumb translator. It solves a number of issues including the necessity for a human translation. Our program allows deaf and dumb individuals to express themselves.

4.1 Advantages

- Project is mainly based on CNN and Media Pipe. We don't use any external device to contact with the system. We can overcome Latency, Robustness and performance issues.

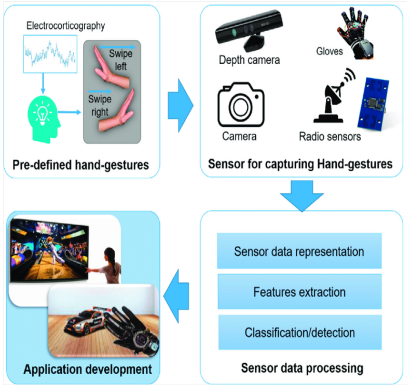
V. SYSTEM DESIGN

5.1 Introduction

System design is the process of defining the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through that system. It is meant to satisfy specific needs and requirements of a business or organization through the engineering of a coherent and well-running system. Systems design mainly concentrates on defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. Systems design implies a systematic approach to the design of a system. It may take a bottom-up or top-down approach, but either way the process is systematic wherein it takes into account all related variables of the system that needs to be created—from the architecture, to the required hardware and software, right down to the data and how it travels and transforms throughout its travel through the system. Systems design then overlaps with systems analysis, systems engineering and systems architecture. The systems design approach first appeared right before World War II, when engineers were trying to solve complex control and communications problems. They needed to be able to standardize their work into a formal discipline with proper methods, especially for new fields like information theory, operations research and computer science in general.

5.2 Dataflow Diagram

A data flow diagram (DFD) illustrates how data is processed by a system in terms of inputs and outputs. As its name indicates its focus is on the flow of information, where data comes from, where it goes and how it gets stored. A data flow diagram maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

Algorithm Description

In Hand gesture Recognition feature extraction and representation is representing an image as a 3D matrix with dimensions such as the height and width of the image and the value of each pixel as depth (1 for grayscale, 3 for RGB). In addition, these pixel values are used to extract useful features using CNN.

6.1Artificial Neural Networks

Artificial neural networks are connections of neurons that mimic the structure of the 19 20 human brain. Each connection in a neuron sends information to another neuron. Input is supplied to the first layer of the neuron, which processes them and sends them to another layer of the neuron called the hidden layer. After processing the information through multiple layers of the hidden layer, the information is passed to the final output layer.

6.2 Unsupervised Learning

Unsupervised learning is a type of machine learning that looks for previously undetected patterns in an existing unlabeled data set and with minimal human supervision. The two most important unsupervised learning methods are principal component analysis and cluster analysis. The only requirement, called unsupervised learning strategy, is to learn a new feature space that captures the properties of the original space by maximizing the objective function or minimizing the loss function. Therefore, the covariance matrix generation is not applicable. Even though the learning is unsupervised, we obtain the eigen symbols ofthe covariance matrix because the eigen analysis operations of linear algebra maximize the variance. This is called principal components analysis.

6.3 Supervised Learning

Supervised learning is a machine learning task that learns a function that maps an input to an output, e.g., an input/output pair. It derives a function from labelled training data consisting of a series of training examples. In supervised learning, each example is a pair of input objects (usually vectors) and desired output values (also known as screens). Supervised learning algorithms analyse the training data and generate derived functions that can be used to map new examples. The best-case scenario allows the algorithm to correctly determine the class labels for the hidden cases. This requires generalizing the learning algorithm.

6.4 Reinforcement Learning

Reinforcement learning (RL) is an area of machine learning. It deals with how software agents behave in their environment to maximize the concept of cumulative reward. Reinforcement learning, along with supervised and unsupervised 22 learning, is one of the three basic paradigms of machine learning. Reinforcement learning differs from supervised learning in that it does not require the presentation of labelled input/output pairs and the need to explicitly modify suboptimal actions. Instead, the focus is on finding a balance between exploration (uncharted territory) and mining (Current knowledge).

6.5 Convolution Neural Networks

Unlike a regular neural network, the neurons in the CNN layer are arranged in three dimensions: width, height, and depth. The neurons in a layer are connected to only a small region of the layer (window size). Before that, instead of all fully connected neurons. Also, at the end of the CNN architecture, the final output layer has dimensions (number of layers) to reduce the overview to a single vector of layer scores The convolution layer uses a small window size (typically 5 * 5 length) that extends to the depth of the input matrix. This layer consists of window-sized learnable filters. During each iteration, we moved the window incrementally [usually 1] to calculate the product of the filter entry and the input value at a particular position. Continue this process to create a two- dimensional activation matrix that reflects the response of this matrix at each spatial location. This means the network will learn A filter that is activated when a visual feature is displayed. B. Edges in a particular direction or spots of a particular color.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

6.6 TensorFlow

TensorFlow is an open-source numerical software library calculation. First define the nodes of the calculation graph, then the actual calculation is done within the session. TensorFlow is widely used in machine learning.

6.7 OpenCV

OpenCV (Open-Source Computer Vision) is an open-source library of programming functions used for real-time computer vision. It is mainly used for analysis of functions such as image processing, video recording, and face and object recognition. Written in C ++, the primary interface, Python, Java, and MATLAB / OCTAVE bindings are available.

VII. CONCLUSION

AI Driven Hand Gesture Recognition through different approaches. We tried different approaches towards AI Driven Hand Gesture Recognition. First, we tried Radar Gesture Recognition with Google Deep Soli Radar Gesture Dataset. Though we achieved good accuracy over there, the radar data failed to recognize the static gestures and then the Radar, has its advantages as it can survive in any climate and it can penetrate through some mediums too. Sine, the Radars were unable to recognize the Static Gestures and Radar Data need lot of computational power. Now, we tried camera approach where we had to recognize the American Sign Language, we trained out 10 digits and then we have included a translator too in the program and then we had good accuracy over there too but we had around 60% validation accuracy over there. Since, it is very time conserving as we had to train a lot of gestures and then we have topre-process the images, the captured images then go through a filter to remove background and then a mask to extract the main features and then the images will be gone through a CNN to classify the gestures over there and then the sentence can be translated using text to speech. The third approach is using the Media Pipe, the Media Pipe recognizes the key points of the hand and then the trained models will be used to detect the hand gestures over there. Due to the disadvantages, we faced while running the model, we believe that adding images of all classes with different backgrounds and distances will cover a wider area and flexibility for live demonstrations. To further improve the model, you need to add a mirrored image as part of the left-handed dataset to perform data expansion to make it translation- invariant equivariant. This prevents the model from over-adjusting certain backgrounds, distances, and hand positions in the image. This allows you to merge the current model with the new dataset to improve the prediction accuracy of each character. Overall, the more images with different characteristics, the better the results. Instead of just recognizing letters, we can try to extend the model to recognize words, phrases, and coherent phrases. Once we have developed a model for generating text/annotations that form logical sentences, we can continue with natural language processing so that we can run various analysis based on the text we extracted. output, such as sentiment analysis to find context. and the feelings behind the words. This is significant because the model will combine two different cognitive computing methods: image recognition and text analysis. In other words, this model will become one of the representative examples of how to extract text from images.

REFERENCES

[1]. R. R, Subramanian, D. Achuth, P.S. Kumar, K. Naveen Kumar Reddy, S. Amara, A. S. Chowdary. (2021). Skin cancer classification using Convolutional neural networks. 11th International Conference on Cloud Computing, Data Science & Engineering.

[2]. R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy, & R. R. Sudarshan. (2022). Airline Fare Prediction Using Machine Learning Algorithms. 4 th International Conference on Smart Systems and Inventive Technology (ICSSIT)

[3]. R. Raja Subramanian, & Karthick Seshadri. (2018). Design and Analysis of Hybrid Hierarchical Feature Tree based Authorship Interface Technique. Advanced in Data and Information Sciences, Springer, 2, 89-104.

[4]. R. Raja Subramanian, & V. Vasudevan. (2021). A deep genetic algorithm for human Activity recognition leveraging fog computing frameworks. Journal of Visual Communication and Image Representation, 77, 1047-3203.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 7.301
Volume 3, Issue 3, April 2023

[5]. R. Raja Subramanian, H. Mohan, A. Mounika Jenny, D. Shrestha, M. Laksh Prasanna & P. Mohan. (2021). PSO Based Fuzzy-Genetic Optimization Technique for Face Recognition. 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

[6]. Raja Subramanian R, & Vasudevan V. (2021). An Ensemble Deep Learning Model For activity Recognition Leveraging IOT and Fog Architectures, In: Gunjan V.K., Zurada J.M. (eds) Modern Approaches in Machine Learning.