International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

Sign Language Identification based on Hand Gestures

Madugula Karthik Sharma[1], Kolusu Mounika[2], Bulusu Satya Sai Bhanu Ruthwik[3], Preethi Jeevan[4]

[1,2,3] B.Tech Scholars, Dept of Computer Science and Engineering Hyderabad-501301,India
[4] Professor, Dept of Computer Science and Engineering, SNIST, Hyderabad-501301, India

---

***Abstract*—Communication is one of the most important part of our day to day life. Inability to speak is considered as a disability. People with this problem often use different modes to communicate with others. One of the major method that is being used is sign language. Our project aims in bridging the gap between normal people, deaf and dumb people using sign language. The main goal is to create a version based system language gestures from the video sequences. The algorithms used in this project mainly consist of CNN, RNN. Video sequences contain both the temporal as well as the spatial features. So we have used two different models to train both the temporal as well as the spatial features. To train the model on the spatial features of the video sequences we have used Inception model which is a deep CNN (convolution neural network).**

Keywords—face-detection, face recognition, feature extraction, harcascade, lbph

I. INTRODUCTION

Movements of body parts such as face and hands gesture. For recognizing gestures we use images here processing and computer vision. gesture recognition, enabling computer and human behavior, Interpreter between computer and human. It also provides the ability to interact with the computer Gestures are performed without physical contact deaf community that uses sign language. The community uses sign language for communication. Speaking is not an option and typing is not a viable option. But there is a possibility of vision. then sign language the only means of exchanging information Man. sign language has the same meaning Spoken words are. It is an isolated type two letters language and continuous sign language. we mainly focus on isolated sign language. Various manufacturers have developed various sign language systems, but they are neither effective nor do they meet the needs of the average person. It's not possible in the average person's budget. Our strategy is to set up an application that uses hand gestures to communicate using predefinedAmerican Sign Language (ASL). Our templates have the latest technology and a wide range of features to meet the needs of the deaf. I used Signs to create an application that can recognize hand gestures. Depending on who you ask, the shape and orientation of your hand will be different. Everyone uses different poses based on their preferences. I need a program that can handle the complexity of the input. A serious problem of non- linearity of hand movements needs to be addressed. Using metadata in addition to data collection can easily solve this problem. In it there is no information about information. Instead of exposing the data itself, the nature of the data stored is disclosed. Gestures can be identified by examining photo metadata and content information. Feature extraction and classification are combined in one operation. Before we can recognize a gesture, we need to get the visual attributes. After extracting such attributes, one needs to apply one of the classification techniques. The main challenge is finding a way to extract core features and apply them to classification. Recognizing and identifying someone requires many traits.

II. LITERATURE SURVEY

Deaf Mute Communication Interpreter: This paper is intended to cover a variety of common method for a deaf communication interpretation system. Two Broad Categories of Communication Methods used by deaf people are wearable communication devices and online learning system. Wearable communication methods include glove-based methods, keyboard methods and touch screen. All the above three split methods use different sensors, accelerometers. A suitable microcontroller, text-to-speech module, keyboard, and touch screen one need can overcome external devices to interpret messages between deaf and non-deaf. The second method, namely online learning systems, there are many different methods of online learning systems. 5 division methods are SLIM module, TESSA, Wi-fi technology, SWI_PELE system, web sign technology

Hand gesture Recognition using PCA: In this paper, the authors presented a scheme using database-driven hand gesture recognition based on the skin color model approach and the threshold approach efficient template matching can be usefully used for human-robot applications, for example. Application: First, the hand region is segmented by applying a skin color model in the YCbCr color space then with a level threshold is applied to separate the foreground and background. Finally, template-based matching the technique is developed using principal component analysis (PCA) for detection.

Hand Gesture Recognition System for Dumb People: The authors presented a static hand gesture recognition

system using digital image processing. The SIFT algorithm is used for hand gesture feature vectors. SIFT features were computed on edges that are invariant to scaling, rotation, and adding noise.

Hand gesture recognition for sign language recognition:

In this paper, a method for automatic character recognition based on shape-based features is presented. Otsu's thresholding algorithm is used for hand region segmentation from the image. This selects the optimal threshold that minimizes the within-class variance of thresholded black and white pixels. The segmented hand region features are computed using Hu invariant moments and fed to an artificial neural network for classification. System performance is evaluated based on accuracy, sensitivity, and specificity.

Sign Pro-An Application Suite for Deaf and Dumb: In this article, the authors proposed a system to help deaf people communicate with the general public using Indian Sign Language (ISL), where the sign language is converted into corresponding text messages. The main goal is to develop an algorithm that converts dynamic gestures to text in real time. Finally, once testing is complete, the system will be implemented on the Android platform and made available as an application for smart phones and tablets.

III Methodology

Most researchers reclassified gesture recognition systems into three main steps Capture input images from devices with cameras, videos, or even data gloves. These here are the steps: Extraction methods, feature estimation and extraction, and classification or recognition as shown in Figure 1.

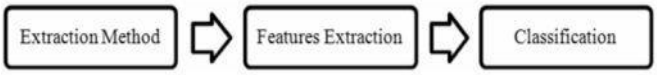

Figure 1. Gesture recognition system steps.

III.I Data Collection

A webcam collects the images needed for model training and validation. Gestures are given by people in front of the webcam. The input image is Use only one hand, gesture with right hand, palm facing camera. The corresponding hand is almost vertical. Makes the recognition process easier the less complex the background and the higher the contrast on the right side the better hand. I'm hoping that the image backgrounds will be simpler and more consistent.

III.II Pre-Processing

Minimal pre processing was applied to the dataset to reduce complexity regarding calculation. It also helps improve efficiency. Use a background Subtraction method, image background removed suggested by Z. ZivKovic. It is mainly based on the K Gaussian distribution which selects the appropriate Gaussian distribution. More adaptable in different scenes with changes distributed to each pixel with lighting. After removing the entire background, only the image of the hand remains. These images have been converted to grayscale. They only contain his one color channel. this will Sign Language Recognition by Hand Gestures 20 It makes CNNs easier to learn and adapt. Morphological erosion was then applied. Then I applied a median filter to reduce the noise. When it comes to signal processing, it only matters Reduce noise.

III.III Features Extraction

A good segmentation process leads to a complete feature extraction process. The latter is important An important role in the successful recognition process. Identifies the vector of the segmented image. It can be extracted in various ways depending on the application. Various methods have been used Applied to represent features that can be extracted. There is also a method using the shape of the hand others used fingertip positions, palm center, etc., such as the outline and silhouette of thehand. 13 parameters were created as feature vectors. The first parameter represents the aspect ratio of the bounding box of the hand, and the remaining 12 parameters are the average value of the brightness pixels within the Photo. When we capture the hand shapeusing the Self-Growing and Self-Organized Neural Gas (SGONG) neural algorithm, we get three features. Palm area, palm center, and hand tilt.

III.IV Gesture classification

After modeling and analyzing the input hand image, the gesture classification method is used Recognize gestures. Appropriate selection of features influences cognitive processes parameters and suitable classification algorithms. For example edge detection or contour. Many hand poses are generated, It can lead to misclassification. Euclidean distance metric used to classify gestures. HMM tools, a statistical tool used for gesture classification, together with finite state machine (FSM) learning vector quantization and principal component analysis (PCA) have demonstrated their ability to recognize dynamic gestures. Neural networks are widely used in the fields of hand shape extraction

[14] and hand gesture recognition. Other soft computing tools such as fuzzy C-means clustering (FCM) and genetic algorithm GA are also effective in this area.

.

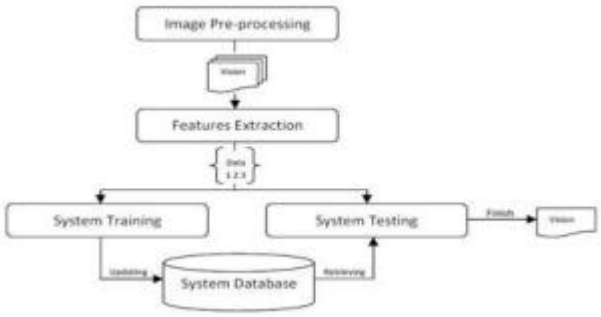

Figure-2 Architecture of gesture recognition system

III.V Data set

We created over 500 different images for each of the gesture provided by ASL(American Sign Language) 37 gestures and for each gesture around 500 different images were taken in all possible angles so that we would receive accurate result.

IV ALGORITHM

Conv-Net/CNN has weights and A dependency that can be explored for its various parts. I can see the difference between the two other classifications. The algorithm requires more preprocessing than ConvNet. Filters are created manually using simple methods, but are not enough.

IV.I Working Model of CNN

After the primitive feature extraction phase, the nextstage Determines the parts of the given object based on the extracted features. Object parts are used to interpret the object's class. A touch-free interface not only increases the driver's concentration and prevents possible breakdowns, Due to its implementation the device ismuch more user friendly such techniques in multiple control systems are preferred.

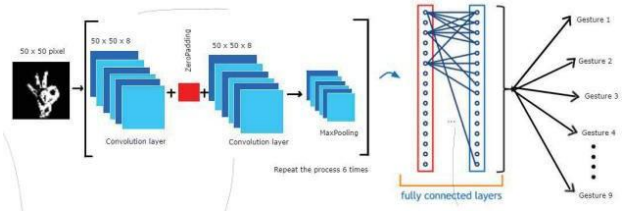

Figure 2: Classification using CNN

IV.II Architecture of CNN

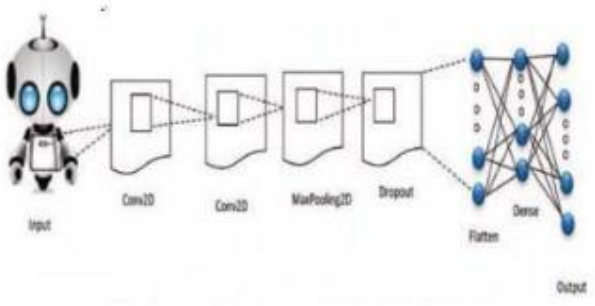

Figure 3 Architecture of CNN

CNN is a type of augmented neural network. specific layers such as

1. Input Layer
2. Hidden Layer
    - 2.1. Convolutional Layer
    - 2.2. A Max Polling Layer
    - 2.3. A Fully Connected Layer
3. Output Layer

Convolutional layer is the first hidden layer in convolutional deep learning and its main purpose is detection and extraction Features from an image that is the input for the model, such as Image edges and vertices. Let's take an example. There is an input image to recognize and it fades out Convolutional layers help the system find edges This picture. Then define the filter matrix, also called the filter matrix .A "kernel" that is further used to find a new image with all edges. Slide this filter across the image to perform specific steps to find a new image with only edges. Apply the dot product to the pixels of the image to find the image that contains them all corner. Edge images are useful for the first steps and layers of CNN. In fact, this is the first set of CNN primitive functions. Works with all feature hierarchies.

Matrix element values are updated optimally value during the training phase. Take the pixel dot product at each matrix value and pan the kernel window. Above the image find a new matrix called the convolved image. As a result, multiple kernels are applied to the image with multiple folded matrices.

Use various kernel help to help find Deviant patterns present in the image, such as curves, edges, etc. At the end of this layer we get a feature map .Output of the folding process. Rectified linear unit(ReLu) .Provides a mapping between response variables and inputs. Swap all negative numbers with '0' and all positive numbers The same is true when using the ReLu function (tf.nn.relu(x)).

For pooling layers, perform dimensionality reduction of activated neurons by down sampling the output image of the relu function. Perform dimensionality reduction using the Max Pooling layer. Max Pooling finds the extremes of the input and simplifies the input. Reduce the number of parameters in your model and transform lower-level data into higher-level information. As shown in figure 4 and figure 5.

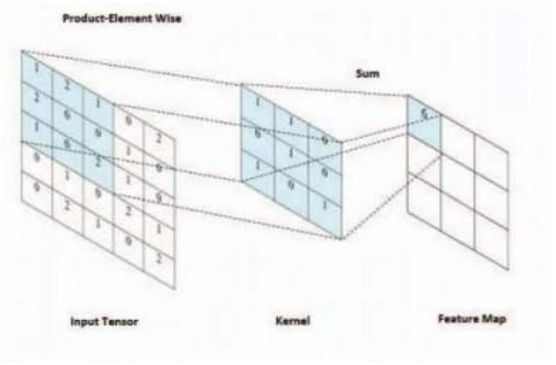

Figure 4 Extracting Fractures

IV.II Classifier

Training, ConvNets can learn these filters and functions. The networks consist of six 2D convolutional layers each followed by the max pooling operator. The output of the sixth convolutional layer is given as the input of a fully connected 9-layer network. Each layer has 512 hidden neurons; except for the last output layer .In this he has 9 neurons, 1 neuron for each of the 9 hand gestures. The output layer uses a sigmoidal activation function. Tongue Activation functions are used in the remaining eight layers. In connection with this article, Each topic takes time, Seeing a real application as a useris unrealistic Often does not stand up to hourly data recording education. To solve this over fitting problem, batch normalization is used and explained in more detail below subsection.

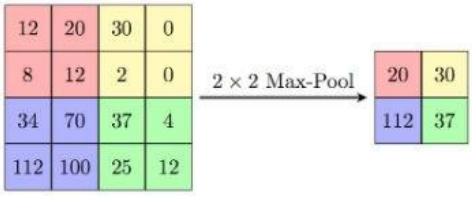

Figure 5 Pooling Function Working

The advantages of CNN include powerful Models that compute faster, are more efficient in terms of memory, and display accurate results in images.

IV.III Activation Function

A modified row execution function (ReLU) is part of a row function that directly takes and returns an input when the input is negative. Zero if the input is positive. Learning is easier and more effective, which is why it is the default enabler for many users. A type of neural network and a modified linear activation function solves the vanishing gradient problem and enables faster model training and work better. For example, a 2x2 matrix becomes a single pixel data by choosing the maximum value of the matrix. direct the steps Sliding behavior of the max pooling process. For example, a stride value of 2 prevents overlap. The window moves 2 pixels each time. Fully connected layers get high-level images from the filtered ones of the previous layers Print it out and convert it to a vector. Each layer of the previous layer.The matrix is first converted to 1D (flattened) dimension For vectors, each vector is fully connected to the next layer connected via the weight matrix. Softmax is an activation function to help find the class of each digit Generate cost probabilities.

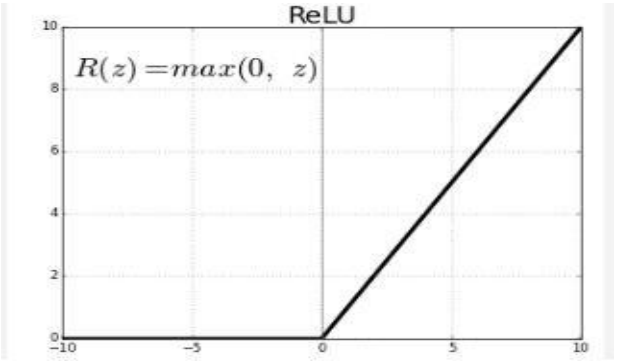

Figure 6 Activation Function (ReLU) GraphV UML Diagrams

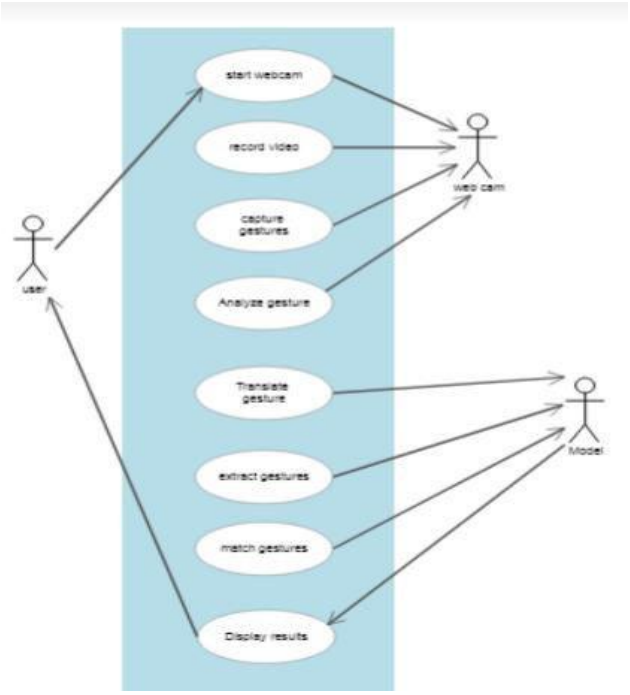

Figure 7 Use case diagram

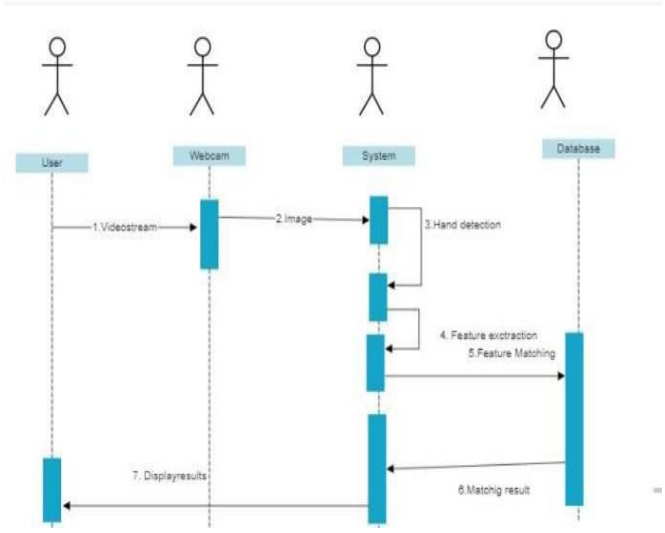

Figure 8 Sequence diagram

VI Implementation

We use the Python programming language to implement and develop this friendly deaf-mute system. Python IDE is Jupiter notebook used for code development and execution. Create a CNN classifier leveraging Keras library features. ToI use Matplotlib for visualization. It contains information such as model accuracy, loss values, and confusion matrix. NumPy was used to perform array operations. Besides the dataset, the training process has two steps.

1) Training with base dataset: The model was trained on the baseline dataset obtained after preprocessing.

2) Advanced dataset training: During this phase, the dataset was expanded.

Data augmentation is a means to increase the number Break up blocks of data using techniques such as zooming, shearing, rotating, and mirroring. This strategy

Not only the data, but also the diversity of datasets required for CNNs to learn differences in localized images.

Input: we will provide a gesture using our hand
Output: It will identify the gesture and display the appropriate symbol

Figure 9 Image to scan the gesture as done

Figure 7 Image displaying done

As shown in figure 8 our model is displaying the correct output as we trained.

We will train our model will all the different alphabets and numbers. Each character i.e. alphabet or number is trained with at least 500 images with all different angles so that we will the results more accurately.

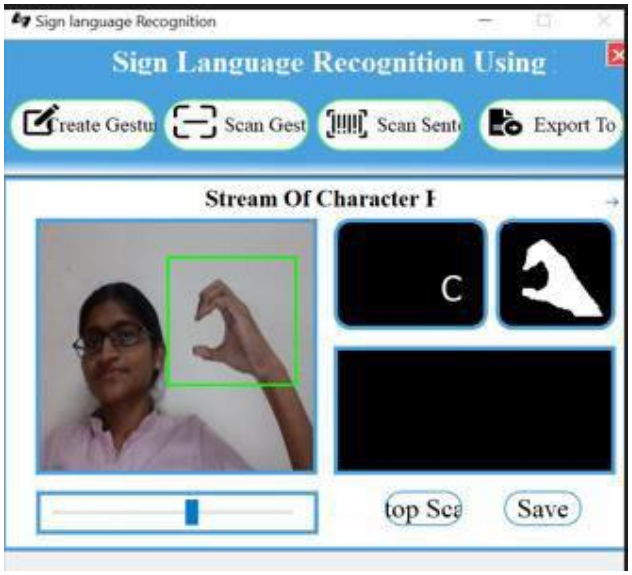

Figure 10 Image to scan the letter c

Figure 11 Image to scan the letter o

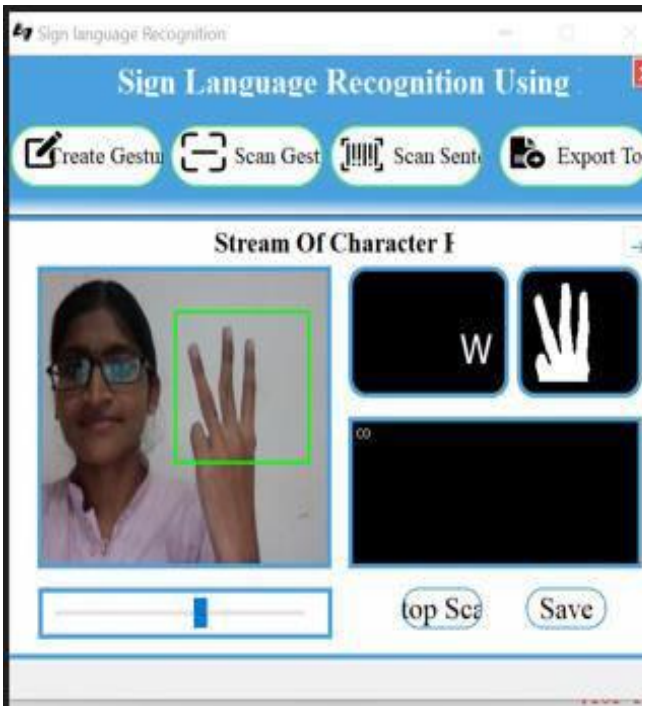

Figure 12 Image to scan the letter w.

Figure 13 Figure displaying output as cow
The output will be displayed as shown in figure 13

VII. Conclusion

In conclusion we can say that communication is an important accepts of daily life. It's hard for people who are dumb and deaf to communicate as normal people. Our project will act as a bridge between normal people and people with such disability helping them to participate in various activities. This project not only used in helping the differently able people, we can also develop a language through which we could encrypt the data so that no one but the desired people could understand what we want to express. We have tested this model on various datasets and have a high accuracy of 95 percent which is very robust as compared to the existing system does not need any predefined templates.

VIII. References

[1] NIDCD, "american signlanguage", 2017 https://www.nidcd.nih.gov/health/americansign- language

[2] Shobhit Agarwal, "What are some problems faced by deaf and dumb people while using todays common tech like phones and PCs", 2017 [Online]. Available: https://www.quora.com/What-are-some- problems-faced-by-deaf-and-dumb-peoplewhileusing- todays-common-tech-like-phones-and-PCs, [Accessed April 06, 2019]

[3] M. Ibrahim, "Sign Language Translation via Image Processing", https://www.kics.edu.pk/project/startup/203

[10] Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 3(2).

[11] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[12] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi: 10.1109/TSMCC.2007.893280.

[13] Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi, Akira Iwata, (2002). "A Rotation Invariant Approach On Static-Gesture Recognition Using Boundary Histograms And Neural International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012

Networks," IEEE Proceedings of the 9th International Conference on Neural Information Processing, Singapura.