SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

[1]M. Nishitha Reddy, [2]G. Rakesh , [3]G. Meghana, [4]B. Niveditha, [5]Dr. R. Jegedeesan, [6]Dr. M Sujatha,[7]Dr.D.Srinivas

IV Year CSE Students[1,2,3,4]Associate Professor[5,6,7]

Jyothishmathi Institute of Technology And Science, Karimnagar

ABSTRACT

Sign language is a method of communication which uses various hand gestures and movements. Understanding these gestures can be postulated as a pattern recognition problem. Humans use different kinds of gestures and motions to convey different messages to other humans. This project represents a framework for a human computer interface capable of recognizing said gestures from sign language and providing a text output representing the meaning of the gesture. The proposed system will use convolutional neural networks and long short term memory networks to identify and learn the gestures which will help to minimize the communication barrier between signers and non-signers.

1. INTRODUCTION

Sign language is a language for the deaf and dumb which uses simultaneous orientation and movement of hand shapes instead of acoustically conveyed sound patterns. Deaf and Dumb people rely on sign language interpreters for communications. However, finding experienced and qualified interpreters for their day to day affairs throughout life period is a very difficult task and also unaffordable. Sign language is the basic means of communication for those with hearing and vocal disabilities. Those disadvantaged have difficulty in their day to day lives.

We aim to develop a system that would ease this difficulty in communication. Sign language consists of making shapes or movements with your hands with respect to the head or other body parts along with certain facial cues. A recognition system would thus have to identify specifically the head and hand orientation or movements, facial expression and even body pose. We propose the design for a basic yet extensible system that is able to recognize static and dynamic gestures of American Sign Language, specifically the letters a-z (where j and z

are dynamic with hand movement while the rest are static). American Sign Language was chosen since it is utilized by a majority of those disabled.

2. RELATED WORK

Over 100 million people - more than 1% of the world's population - are unable to hear. Being deaf from birth or childhood, many of these people use sign language as their primary form of communication. There is an undeniable communication problem between the Deaf community and the hearing majority. Innovations in automatic sign language recognition try to tear down this communication barrier. Our contribution considers a recognition system using the convolutional neural networks (CNNs) and GPU acceleration. Instead of constructing complex handcrafted features, CNNs are able to automate the process of feature construction.

We aim to reduce the gap that exists between signers and non-signers by using Convolutional Neural Networks (CNNs). We first extract features from the frame sequences resulting in a representation consisting of one or more feature vectors. The process will be carried by a convolutional neural network trained on the reduced dataset. This aims the computer to differentiate between possible classes of actions. The second step is video classification of the gestures. This is done by Long Short Term Memory (LSTM), a type of Recurrent Neural Network (RNN).

3. METHODS

The system is a vision-based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

A. DATASET GENERATION

It is required to make a proper database of the gestures of the sign language so that the images captured while communicating using this system can be compared. Steps we followed to create our data set are as follows. We used Open computer vision (OpenCV) library in order to produce our dataset. Firstly, we captured around 800 images of each of the symbol in ASL for training purposes and around 200 images per symbol for testing purpose. First, we capture each frame shown by the webcam of our machine. In each frame we define a region of interest (ROI) which is denoted by a blue bounded square as shown in the image below. From the whole image we extracted our ROI which is RGB and convert it into grey scale

Image. Finally, we apply our Gaussian blur filter to our image which helps us extracting various features of our image.

B. GESTURE CLASSIFICATION

The approach which we used for this project is Our approach uses two layers of algorithm to predict the final symbol of the user.

Algorithm Layer 1:

1. Apply Gaussian blur filter and threshold to the frame taken with opencv to get the processed image after feature extraction.

2. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for
Forming the word.

3. Space between the words are considered using the blank symbol.

Algorithm Layer 2:

We detect various sets of symbols which show similar results on getting detected. 2. We then classify between those sets using classifiers made for those sets only.

C. TRAINING AND TESTING

We convert our input images (RGB) into grayscale and apply Gaussian blur to remove unnecessary noise. We apply adaptive threshold to extract our hand from the background and resize our images to 128 x 128. We feed the input images after preprocessing to our model for training and testing after applying all the operations mentioned above.

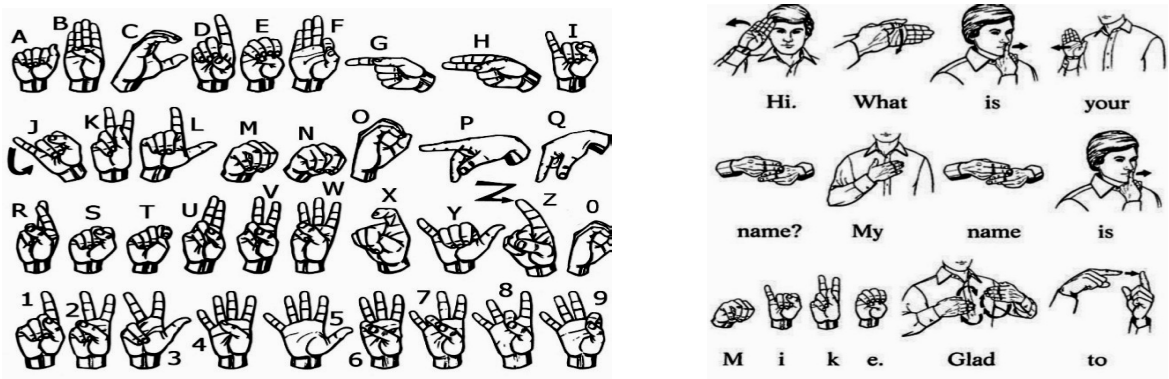

Fig: sample datasets for sign language

4. SYSTEM DESIGN

Image Acquisition:

It is the action of extracting an image from a source, typically a hardware-based source, for process of image processing. Web Camera is the hardware-based source in our project. It is the first step in the workflow sequence because no processing can be done without an image. The picture that is obtained has not been processed in any way.

Segmentation:

The method of separating objects or signs from the context of a captured image is known as segmentation. Context subtracting, skin-color detection, and edge detection are all used in the segmentation process. The motion and location of the hand must be detected and segmented in order to recognise gestures.

Features Extraction:

Predefined features such as form, contour, geometrical feature (position, angle, distance, etc.), colour feature, histogram, and others are extracted from the preprocessed images and used later for sign classification or recognition. Feature extraction is a step in the dimensionality reduction process that divides and organises a large collection of raw data. Reduced to smaller, easier-to-manage classes As a result, processing would be simpler. The fact that these massive data sets have a large number of variables is the most important feature. To process these variables, a large amount of computational power is needed. As a result, function extraction aids in the extraction of the best feature from large data sets by selecting and combining variables into functions. reducing the size of the data These features are simple to use while still accurately and uniquely describing the actual data collection.

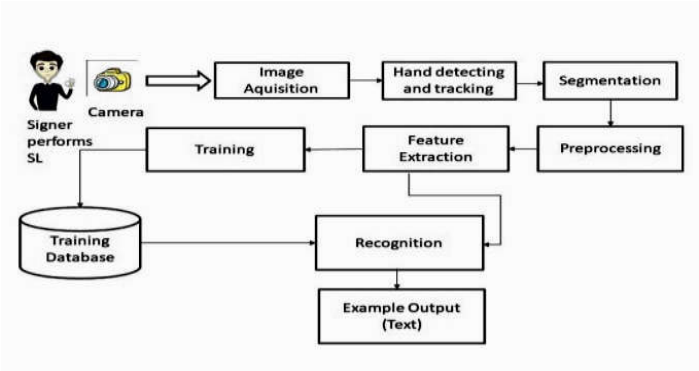

Fig: system design

Pre-processing:

Each picture frame is preprocessed to eliminate noise using a variety of filters including erosion, dilation, and Gaussian smoothing, among others. The size of an image is reduced when a colour image is transformed to grayscale. A common method for reducing the amount of data to be processed is to convert an image to grey scale.The phases of preprocessing are as follows:

Morphological operations use a structuring feature on an input image to create a similar-sized output image. It compares the corresponding pixel in the input image with its neighbours to determine the value of each pixel in the output image. There are two different kinds of morphological transformations Erosion and Dilation.

Recognition:

We'll use classifiers in this case. Classifiers are the methods or algorithms that are used to interpret the signals. Popular classifiers that identify or understand sign language include the Hidden Markov Model (HMM), KNearestNeighbor classifiers, Support Vector Machine (SVM), Artificial Neural Network (ANN), and Principle Component Analysis (PCA), among others. However, in this project, the classifier will be CNN. Because of its high precision, CNNs are used for image classification and recognition. The CNN uses a hierarchical model that builds a network, similar to a funnel, and then outputs a fully-connected layer in which all neurons are connected to each other and the output is processed.

Text output:

Understanding human behaviour and identifying various postures and body movements, as well as translating them into text.

5. ANALYSIS

5.1 EXPERIMENTATION

The CNN is tested with the dataset that has been generated. A total of 1,03,000 datasets comprising 2400 samples for 43 gestures is used for classification and training. The Dataset is derived from a Finger Spelling Library, which is a open source project for collection of Sign Language Datasets.The user makes the gestures in-front of the camera which captures the Depth Image and separates the gestures from the entire scene.

The important features such as position of hand and fingers are extracted to make decision on what is being gestured.

The Gesture Recognition is performed using Convolutional Neural Networks, by classifying the gestures. The sentences are interpreted from the gesture and finally sent to the user interface which displays the sentence.

5.2 RESULTS

Screen shots:

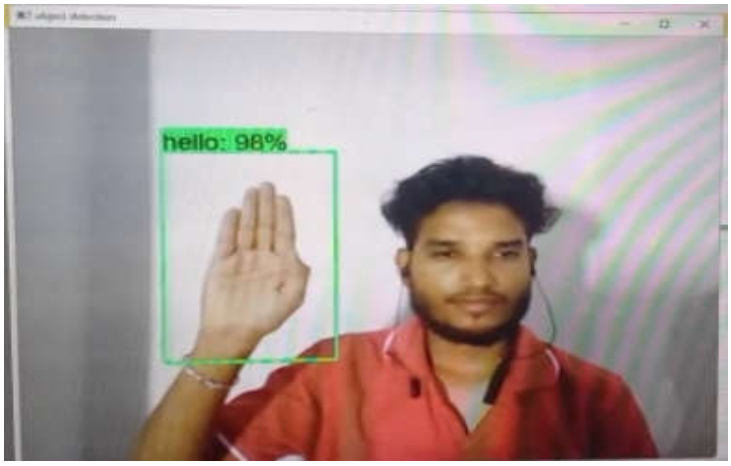

Fig: A sign of hello

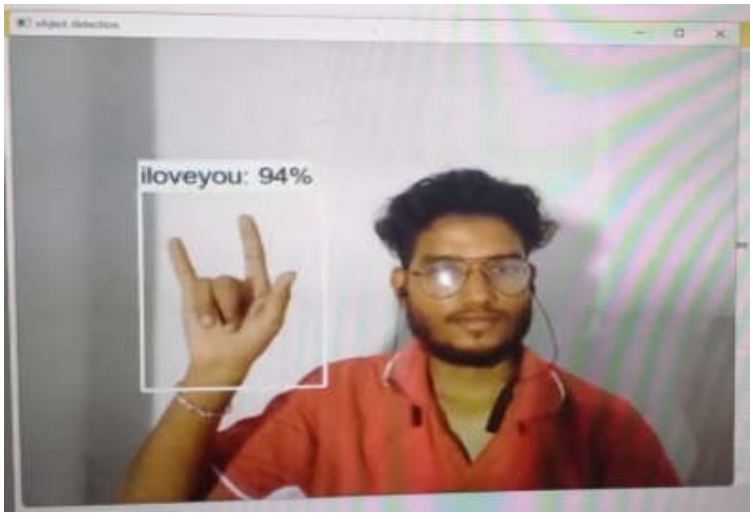

Fig: A sign of i love you

Fig: A sign of no

6. SUMMARY

6.1 CONCLUSION

Many breakthroughs have been made in the field of artificial intelligence, machine learning and computer vision. They have immensely contributed in how we perceive things around us and improve the way in which we apply their techniques in our everyday lives. Many researches have been conducted on sign gesture recognition using different techniques like ANN, LSTM and 3D CNN. However, most of them require extra computing power. On the other hand, our research paper requires low computing power and gives a remarkable accuracy of above 90%. In our research, we proposed to normalise and rescale our images to 64 pixels in order to extract features (binary pixels) and make the system more robust. We use CNN to classify the alphabetical sign gestures and successfully achieve an accuracy of 98% which is better than other related work stated in this paper.

6.2 FUTURE SCOPE

This project holds immense potential in terms of real world applications and can be used as a platform for development of solutions to a number of problems. The future scope of this project includes increasing accuracy of current system: This is done in two ways. Firstly, by increasing the size of training dataset to include more variations of characters. Accuracy can also be improved by training the data to a deeper level. However, this will require a very high configuration machine.

7. REFERENCES

[1] Goodman J W 1968 Introduction to Fourier optics McGraw Hill

[2] N.Otsu, “A Threshold Selection Method from Gray-LevelHistograms”, IEEE transactions on systems, man, and

[3] Rafiqul Zaman Khan and Noor Adnan Ibraheem , “HandGesture Recognition: A Literature Review”, InternationalJournal of Artificial Intelligence & Applications (IJAIA),Vol. 3, No. 4, pp. 162-174, July 2012

[4] M. Panwar, "Hand gesture based interface for aiding visuallyimpaired," Proc. IEEE Int. Conf. Recent Adv. Comput. Softw.Syst. RACSS2012, pp. 80-85.

[5] Aliaa A. A.Youssif, AmalElsayedAboutabl, HebaHamdyAli,"Arabic Sign Language (ArSL) Recognition System UsingHMM ", (lJACSA) International Journal of AdvancedComputer Science and Applications, Vo1.2, Issue. 11,2011

[6] L. Gu, X. Yuan, and T. Ikenaga, "Hand gesture interface basedon improved adaptive hand area detection and contoursignature," IEEE Int. Symp. Intel!. Signal Process. Commun.Syst. (ISPACS 2012), no. Ispacs, pp. 463—468

[7] H. Y. Lai and H. J. Lai, "Real-Time Dynamic Hand GestureRecognition," IEEE Int. Symp. Comput. Consum. Control,2014no. 1,pp. 658-661

[8] Maeda Y, Wakamura M 2005 Simultaneous perturbation learning rule for recurrent neural networks and its FPGA implementation IEEE Trans. Neural Network 16 6 1664 – 1672.

[9] Gallaudet University Research Institute: A Brief Summary of Estimates for the Size of the Deaf Population

[10] Dr.A.Nirmal Kumar, Dr.R.Jegadeesan, Dr.D.Baswaraj, J.Greeda 2019 “Improved Migration Performance In Virtualized Cloud Datacenters” International Journal of Scientific & Technology Research. Volume 8, Issue 09,page no.1515-1518 September 2019. (Scopus indexed)

[11].Annadi Jahnavi, Hanumandla Bhavana, Dr. R. Jegadeesan “An Implementation of Detecting Password Pattern In Dictionary Attack” International Journal of Advanced Science and technology. ISSN 2005-42386, Page no: 84-92 July,2019(Indexed by Scopus, Elsevier).

[12].Dr.A.Nirmal Kumar, Dr.R.Jegadeesan, Dr.C.N.Ravi, J.Greeda2019 “A Secure Transaction Authentication Scheme using Blockchain based on IOT” International Journal of Scientific & Technology Research. VOLUME 8, ISSUE 10,Page no:2217-2221 OCTOBER 2019. ISSN 2277-8616 (Scopus indexed)

[13].Dr R Jegadeesan, Dr.C.N.Ravi, Dr.A.Nirmal Kumar 2020 “Automatic Rice Quality Detection Using Morphological and Edge Detection Techniques” ICCCE 2020 3rd International Conference on Communications and cyber Physical Engineering, Metadata of the chapter that will be visualized in Springer Link. Volume, issue, May.2020 Page No.233-242, Springer conference.