ISSN No:-2456-2165

Si-Lang Translator with Image Processing

Ms. Livya George Asst. Professor
Dept. of Computer Science and Engineering
SCET, Thrissur

Nimisha Elangikkal, Nina Joseph, Ronica Ross, Sharon Joy
UG Students
Dept. of Computer Science and Engineering
SCET, Thrissur

Abstract:- People having hearing and speaking disabilities will have problems communicating with other people. This creates a gap between them. To avoid this problem, they use some special gestures to express their thoughts. These gestures have different meanings. They are defined as "Sign Language". Sign language is very important for deaf and mute people because they are the primary means of communication between both normal people and between themselves. It is most commonly used for people with talking and hearing disorders to communicate. In this application, we present a simple structure for sign language recognition. Our system involves implementing such an application that detects predefined signs through hand gestures. For the detection of gestures, we use a basic level of hardware components like a camera, and interfacing is needed. Our system would be a comprehensive User-friendly Based application built on Convolutional Neural Networks. The hand gestures are recognized by three main steps. First, the dataset is created by capturing images and these images are preprocessed by resizing, masking, and converting RGB into grayscale images. Secondly, after creating the dataset, we have to train the system using the Convolutional Neural Network, and using the trained classifier model the given sign is recognized. Thus, the recognized sign is displayed. We expect that the overall method of this application may attract technophiles with an extensive introduction in the sector of automated gesture and sign language recognition, and may help in future works in these fields.

Keywords:- *Convolutional Neural Network ; Preprocessing ; Sign Language ; ReLU.*

I. INTRODUCTION

Around 20-45% of people all over the world suffer with hearing and speaking disabilities, and about eight per 20,000 of the global population become deaf and dump before learning any of the language. This creates them to start their own countries' sign language as their foremost means of communication. According to topical data of the World Federation of the Deaf, there are over 70 million sign language users in the world, with over 300 sign languages across the globe. Contrary to common thoughts, not all people who speak with the help of sign languages are able to read and understand the normal texts as non-disabled people who use them, which is due to the differences in sentence syntax and grammar.

Sign Language is the means of visual communication, where they use expressions, hand gestures, and body movements as the means of communication. Sign language is significant for people who suffer from difficulty with hearing or speech. Sign Language Recognition refers to the transformation of words or alphabets into gestures into normally spoken languages of their own locality. Thus, the transformation of sign language into words or alphabets can help to overcome the gap existing between impaired people and the rest of people around the world.

A. PROBLEM DOMAIN

Existing systems deal with many problems. ASL alphabet recognition is a challenging task due to the difficulties in hand segmentation and the appearance of the variations among signers. The color-based systems also suffer from many challenges such as complex background, hand segmentation, large inter-class and intra-class variations. These all mechanisms have a practical limitation because it is necessary to use costly extra hardware for getting data for sign recognition.

The existing dynamic sign language recognition methods still have some drawbacks with difficulties of recognizing complex hand gestures, low recognition accuracy for most dynamic sign language recognition, and potential problems in larger video sequence data training. The static sign language recognition is hard to deal with the complexity and large variations of vocabulary set in hand actions. So, it may make a misunderstanding of some significant variations from signers. Dynamic sign language recognition also has challenges in dealing with the complexity of the sign activities of finger motion from the large-scale body background. Another difficulty facing is the extracting of the most discriminating features from images or videos. Besides, how to choose an appropriate classifier is also a critical factor for producing accurate recognition results.

Over a period of time, in the marketplace, there are a variety of products that are capable of converting the signs to text. For example, a wearable glove is able to translate the Indian Sign Language, but the device would need to recognize both the static and dynamic gestures in ISL. While standard neural networks can classify static gestures, they cannot be used for classifying dynamic gestures. In dynamic gestures, the reading at each time point is dependent on the previous readings resulting in sequential data. Since standard neural networks require that each reading be independent of the other readings they cannot be used for the classification of sequential data. Under the exact context of symbolic expressions of deaf and dumb people is a challenging job in the real life unless it is properly specified.

B. PROPOSED SYSTEM

Communication is a big part of everyone's day-to-day life. People with the inability to speak use different modes to communicate with others, one such widely used method of communication is sign language. Developing sign language translation applications for deaf people can be very beneficial, as they will be able to communicate easily with even those who don't understand sign language. Our project aims at taking the basic step in interlacing the communication gap between normal people and dumb people through an easy sign language recognizer.

Sign language translation has always been an engrossing area in Machine Learning. Sign languages have multiple articulators like hands, shoulders, or even parts of the face. Due to these facts, none of the sign language translation techniques can have an accuracy of 100%. But the need for them makes this field a highly researched area at all times. by the success of deep learning technology, it has proven to have a higher recognition accuracy than the traditional methods. From many surveys, it has been concluded that Deep learning can be a better solution for many of the sign language detection shortcomings. By considering several other facts like convenience and cost-effectiveness, methods that are based on CNN serve better for sign language interpretations.

Sign language translation based on CNN methods is basically about training the model with a sign language dataset and thus creating a classifier model that detects the signs. Different countries have various sign gestures. Sign gestures can be either hand gestures alone or a combination of facial expressions and hand gestures. It is difficult to segment the face and hand in cases where the background is noisy. And also differentiating facial expressions is another big task. Hand gestures can be performed either by one hand or two hands. In some situations, the dual hand systems can make confusions and affect predictions. Signs are of two types: Isolated sign language and Continuous sign language. Hand tracking is a challenging task in the continuous model. Considering all these in this project, we will use an isolated one-hand gesture recognition technique and a new self-made gesture dataset is created for the proposed system. In this project, the recognition of hand gestures is done with the usage of Convolution Neural Network.

Using CNN is well-liked due to three important factors:

- The features are learned directly by CNN.
- CNN enables building on pre-existing networks.
- CNN produces highly accurate recognition results.

This approach gets images from a good quality camera and pre-processing steps are performed for the images. These images are given to a convolution 2D network for feature extraction. Based on the features extracted the Conv2D network is trained. Thus, we get a trained classifier model that distinguishes different sign gestures. Now the system can be used for detecting various signs shown by signers. Respective words are displayed on the screen for each sign gesture. The detailed explanations of the system are given in the coming section of this paper. Section 2 shows the overall block diagram of the proposed system and Section 3 deals with the detailed working of the system.

II. SYSYTEM ARCHETECURE

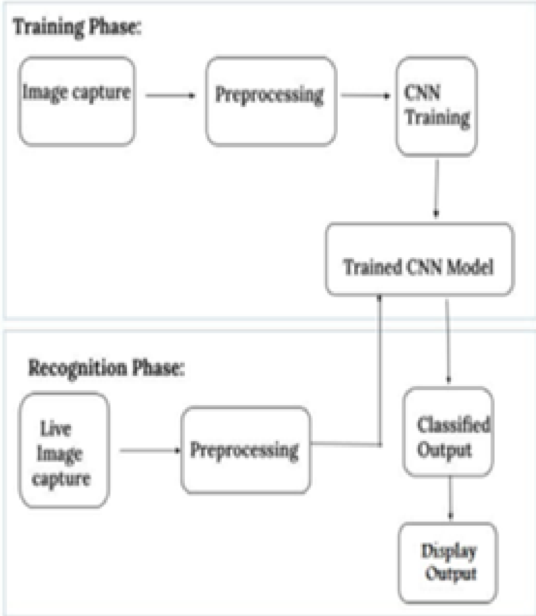

Fig. 1. Block diagram

In the training phase sign images are captured and pre-processed. Its output is given as the input for CNN training and thus a classification model for sign recognition is generated. In the recognition phase, the user shows signs in front of the camera from which the images are captured and pre-processed, and given to the trained classifier model. Thus, predictions are made and the result is displayed to the user on the monitor.

III. WORKING EXPLANATION

The overall process can be classified into four different phases:

A. Dataset creation phase
B. Training phase
C. Sign recognition phase

A. Dataset creation phase

Image capturing is an integral part of the image processing system. For this, we are using a good quality webcam. First, we need to import all needed modules for accessing the webcam and capturing gestures. We should give a label for each sign and then the sign capturing starts. The captured images are given for a pre-processing step. This is done by resizing, masking, and color conversion. Separate image folders are created for training and testing. For each sign, we captured 400 images of which the first 350 images are stored in the training set and the remaining 50 stored in testing.

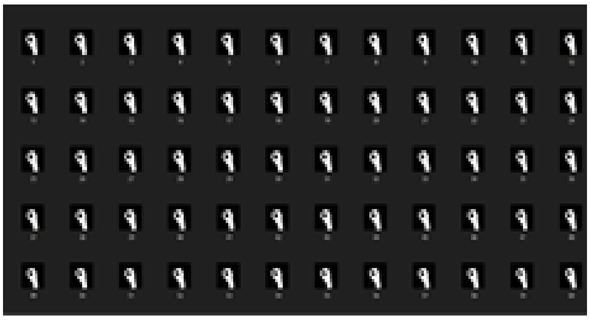

Fig. 2. Data created for the sign "food"

B. Training phase

After the dataset creation, we need to train the dataset on the convolutional neural network model. For this, we import Keras libraries and packages. The procedure for building a convolutional neural network (CNN) involves four major steps.

Step - 1: Convolution
Step - 2: Pooling
Step - 3: Flattening
Step - 4: Full connection

- Convolution: This is the first layer that is used to extract several features from the input images. We are using a sequential neural network model that contains a sequence of layers. Since we are working with sign images which is a 2D array, here we are using Convolution 2D with 32 numbers of 3×3 filters and color images of 64×64 resolution. The activation function we used here is rectified linear activation function (ReLU).
- Pooling: This is mainly done to downsize the images. We used max pooling with a window size of 2×2.
- Flattening: We need to convert the pooled images into a continuous vector form. So, we are using a flattening function. What happens is that the images in a two-dimensional array form are converted into a one-dimensional vector form.
- Full connection: The result after the flattening step is given as the input to this fully connected layer. These layers are placed before the output layer. The classification process begins to take place at this stage. There is a chance for overfitting in the training dataset. Overfitting takes place when a model works so accurately on the training data which reduces the performance of the system when tested with new data. So, a dropout of 0.5 is used here. An activation function of SoftMax is used in this case as it involves multi-class classification.

The training dataset is given to the CNN model for training with an epoch of 25 and the steps in each epoch are 800. The test dataset is validated then. Now the classifier model is ready to recognize different sign gestures.

C. Sign recognition phase

Now the trained classifier model is loaded and Images are captured live. These images are preprocessed and saved. These images are given for prediction. There the image is converted to an array for comparison. Sign with the highest match is recognized. The respective sign label is then returned.

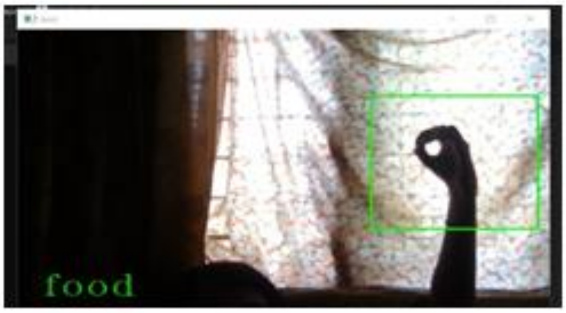

Fig. 3. Recognition of the sign "food"

IV. HARDWARE REQUIREMENTS

- Processor: i5
- RAM: 8 GB
- Hard Disk: 500 GB

V. SOFTWARE REQUIREMENTS

- Python

Python is a general-purpose programming language that may be used for a variety of tasks, including back-end development, software development, and writing system scripts. Python 3.9 was used in this project.

- OpenCV

Open-Source Computer is an open-source computer vision and machine learning software library that is used to provide a common infrastructure for computer vision.

- TensorFlow

TensorFlow is a free open-source software library that is used in machine learning. It is a math library based on data flow and differentiable programming.

- Keras

OpenPose Keras is an open-source software library that gives a python interface for the TensorFlow library.

VI. CONCLUSION

People with speaking and hearing disabilities face much indignation and discouragement that limit their ability to do day-to-day tasks. It is proven that dumb and deaf people, especially impaired children, have a greater chance of behavioral and emotional disorder in the way others discriminate against them. This causes people with such disabilities to become introverts and resist social connectivity and face-to-face communications. Unable to communicate with family and friends can affect their self-esteem and this leads to isolating deaf and dumb people from society. Due to this, they lack social interactions, and also communication skill is also a huge barrier for the deaf and dumb. From this application, we have tried to conquer some of the major problems faced by disabled persons in terms of talking. The result was that people on the other side are not able to communicate what these persons are trying to say or what is the message that they want to convey.

Our application will help those who want to learn and communicate in sign languages. With this, a person will quickly adapt various gestures and their meaning as per the predefined signs. They can quickly learn which gesture to use for each sign. A user need not be a literate person if they know the action of the gesture, they can quickly form the gesture and an appropriate assigned character will be shown onto the screen. These predefined gestures are formed by training the system with CNN. Firstly, a dataset is created. The images of gestures are captured.

These captured images are preprocessed. About 400 images are taken for each gesture in which 350 images are for the training set and the rest 50 are for the test set. After forming the dataset, the training is done for recognizing the gestures. The training is done through the Convolutional Neural Network. In CNN, the process is run through mainly four layers: 1.convolution, 2.pooling, 3.flattening and 4.full connection. After training the gestures, the trained classifier model is loaded. Images captured as live will undergo resizing, masking. These images are then given for prediction. These images are converted to an array. Thus, the sign with the highest match is recognized. Finally, the output will be displayed. As a result, the suggested algorithm is robust. Experiments have proven that the procedure works and that it produces the desired result, as well as a proper alert.

ACKNOWLEDGMENT

We want to offer our heartfelt gratitude and heartfelt thanks to everyone who helped us to make this initiative a huge success. We thank the almighty God for all the benefits he has granted us. Prof. Livya George, our guide, deserves huge thanks. We want to thank our Principal, Dr. Nixon Kuruvila, and our Department Head, Dr. M Rajeswari, for their unwavering support throughout the project.

REFERENCES

[1]. C. Dong, M. C. Leu, and Z. Yin, “American Sign Language alphabet recognition using Microsoft Kinect,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), June. 2015, pp. 44–52.

[2]. J.T.E Richardson, L. Barnes, and J. Fleming,“Approaches to studying and perceptions of academic quality in deaf and hearing students in higher education,” Deafness Educ. Int., vol. 6, no. 2, pp. 100–122, 2004.

[3]. S. Riddell and E. Weedon, “Disabled students in higher education: Discourses of disability and the negotiation of identity,” Int. J. Educ. Res., vol. 63, pp. 38–46, Dec. 2014.

[4]. S. C. Daroque, “Alunos surdos no ensino superior: Uma discussão necessária,” Comunicações, vol. 19, no. 2, pp. 23–32, Jul./Dec. 2012

[5]. R. Kheir and T. Way, “Inclusion of deaf students in computer science classes using real-time speech transcription,” in Proc. 12th Annu. Conf. Innov. Technol. Comput. Sci. Educ. (ITiCSE), 2007, pp. 261–265.

[6]. T. Oliveira, P. Escudeiro, N. Escudeiro, E. Rocha, and F. M.Barbosa, “Automatic sign language translation to improve communication,” in Proc. IEEE Global Eng. Educ. Conf. (EDUCON), Apr. 2019, pp. 937–942

[7]. P. Escudeiro et al., “Virtual sign—A real time bi-directional translator of portuguese sign language,” Procedia Comput. Sci., vol. 67, pp. 252–262, Sep. 2015.

[8]. J. Ulisses, T. Oliveira, P. M. Escudeiro, N. Escudeiro, and F.M.Barbosa, “ACE assisted communication for education: Architecture to support blind & deaf communication,” in Proc. IEEE Global Eng. Educ. Conf. (EDUCON), Apr. 2018, pp. 1015–1023.

[9]. T. Reuters. (2018). Thomson Reuters Web of Science, Web of Science. Available:http://apps.webofknowledge.com

[10]. M. Taskiran, M. Killioglu, and N. Kahraman, “A real-time system for recognition of American sign language by using deep learning,” in Proc. 41st Int. Conf. Telecommun. Signal Process. (TSP), Jul. 2018, pp. 1–5.

[11]. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., June. 2014, pp. 1725–1732. athindello..

[12]. E. K. Kumar, P. V. V. Kishore, A. S. C. S. Sastry, M. T. K. Kumar, and D. A. Kumar, “Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps,” IEEE Signal Process. Lett., vol. 25, no. 5, pp. 645–649, May 2018.

[13]. Y. Liao, P. Xiong, W. Min, W. Min, and J. Lu, “Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks,” IEEE Access, vol. 7, pp. 38044–38054, 2019.

[14]. T. Yuan et al., “Large scale sign language interpretation,” in Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2019, pp. 1–5.

[15]. J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, “Video-based sign language recognition without temporal segmentation,” in Proc. 32nd AAAI Conf. Artif. Intell., 2018.

[16]. G. A. Rao, P. V. V. Kishore, A. S. C. S. Sastry, D. A. Kumar, and E. K. Kumar, “Selfie continuous sign language recognition with neural network classifier,” in Proc. 2nd Int. Conf. Micro-Electron., Electromagn. Telecommun., vol. 434. Singapore: Springer, Sep. 2018, pp. 31–40.

[17]. J. Fellinger, D. Holzinger, and R. Pollard, “Mental health of deaf people,” Lancet, vol. 379, no. 9820, pp. 1037–1044, 2012. [2] (2019). Word Federation of the Deaf, Our Work. Accessed: Sep. 13, 2019. [Online]. Available: http://wfdeaf.org/ our-work/

[18]. Z. Zhang, “Microsoft Kinect sensor and its effect,” IEEE MultiMedia, vol. 19, no. 2, pp. 4–10, Feb. 2012

[19]. P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA, USA: Addison-Wesley, 2005.

[20]. M. Mohandes, M. Deriche, and J. Liu, “Image-based and sensor-based approaches to arabic sign language recognition,” IEEE Trans. Human Mach. Syst., vol. 44, no. 4, pp. 551–557, Aug. 2014.