

Malayalam Sign Language Recognizer

Ajmi N¹, Akhila Raj B R¹, Basina B¹, Varsha S¹, Vishnu S Kumar²

¹UG Scholar, Department of Computer Science and Engineering ,

²Asst.prof, Department of Computer Science and Engineering,
UKF College of Engineering and Technology, Kerala, India

Abstract:- In recent years, human computer interaction become the regular part of our life. The hand gestures are the foundation of sign language, which is a visual form of communication. In this paper, we have designed a robust hand gesture recognition system which can efficiently track static hand gestures. We have created a desktop application that uses a computer's webcam to capture a person signing gestures for Malayalam sign language, and translate it into corresponding text and audio in real time. In this paper, Inception modules are used in Convolutional Neural Networks to allow for more efficient computation. First, in this work, Malayalam sign language consisting of 2000 images for each gesture is collected using RGB camera. the highest accuracy is obtained by the proposed model.

Keywords :- Human-computer interaction ,Hand gesture recognition, Malayalam sign language recognition, Inception

INTRODUCTION

Interaction between humans and computers HCI is a multidisciplinary field of study that focuses on the design of computer technology and, in particular, the human-computer interface.

In recent years, many researchers have been interested in the field of gesture based HCI and gesture recognition. It is the skill of computer to identify hand gesture from sources like images or video feed. Hand gesture recognition is mostly used in the field of sign language recognition. Sign language bridges the gap of communication between deaf people and normal community. The development of effective local level sign language recognition tool is essential.

According to a new estimate from the World Health Organization (WHO), one in every four people, or roughly 2.5 billion people, will have mild-to-profound hearing loss by 2050. According to WHO, at least 700 million individuals will be affected by debilitating hearing loss and will require ear and hearing care.

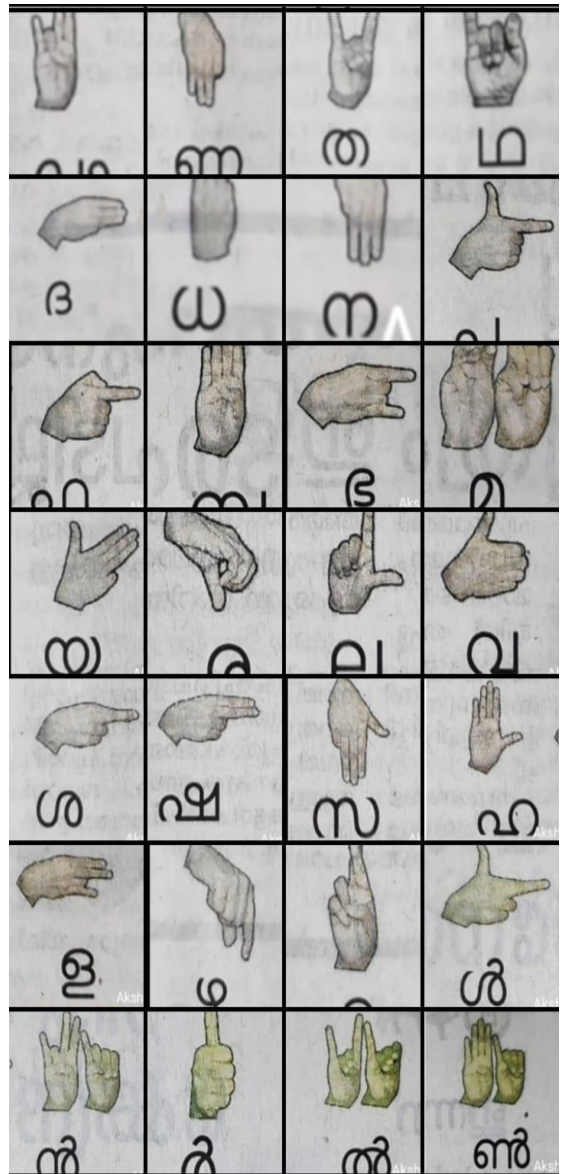
Human-machine interaction, sign language, game technology, robotics, and other fields use gesture recognition. Hand gestures used in sign language can be of two types i.e. static and dynamic gestures. . A static gesture is a single image that depicts a specific hand configuration and pose.

Waving hand, fist hand, vertical hand, and horizontal hand are some of the dynamic hand gestures that have been defined. A dynamic gesture is one that moves and is represented by a series of images. The term "dynamic hand gesture" refers to the recognition of gestures with a dynamic hand. The goal of the proposed research is to create a system that can recognize static sign gestures and convert them into text and audio. To obtain data from the signer, a vision-based approach using a web camera is introduced.

The system was created with the intention of serving as a learning tool for those interested in learning more about the fundamentals of sign language, such as alphabets and common static signs. The proponents provided a white background and a specific location for image processing of the hand, thus improving the accuracy of the system we have used inception model as the recognizer of the system.

RELATED WORKS

In paper [1] demonstrates, Creating a desktop application that uses a computer's webcam to capture a person signing movements for American sign language (ASL) and translate them into appropriate text and speech in real time. The sign language gestures will be translated into text, which will then be turned to audio. They created a finger spelling sign language translator in this way. Convolutional neural networks were employed to enable the detection of gestures (CNN). After appropriate training, a CNN is capable of recognizing the desired features with a high degree of accuracy and is highly efficient in addressing computer vision problems.



It is possible to obtain a finger spelling sign language translator with a 95% accuracy rate.

In [2] Using Support Vector Machine (SVM) and Convolutional Neural Networks, the authors proposed American Sign Language (CNN). They also estimated the ideal filter size for single and double layer Convolutional Neural Networks in this study. The first phase involves extracting features from the dataset. On the training dataset, Support Vector Machines with four different kernels and Convolutional Neural Networks with single and double layers are used to train the model after various preprocessing approaches have been employed.

For both procedures, accuracy is calculated and compared. CNN filters of various sizes were used, and the best filter size was discovered. The accuracy of the single layer CNN is 97.344 % and the accuracy of the double layer CNN is 98.581 %.

In [3] authors used a multimodal dynamic sign language recognition approach called BLSTM-3D residual network, which is based on a deep 3-dimensional residual ConvNet and bi-directional LSTM networks (B3D ResNet). There are three primary aspects to this strategy. To reduce the time and space complexity of network calculation, the hand object is first localized in the video frames.

After feature analysis, the B3D ResNet automatically extracts spatiotemporal characteristics from video sequences and assigns an intermediate score to each action in the video sequence.

Finally, the dynamic sign language is accurately detected by classifying the video sequences. The experiment uses test datasets such as the DEVISIGN D dataset and the SLR Dataset. On the DEVISIGN D dataset, achieve state-of-the-art recognition accuracy of 89.8% and 86.9%, respectively.

In [4], they covered most of the currently known approaches for SLR tasks based on deep neural networks developed over the last several years and classified them into clusters based on their main characteristics. The relative strengths and limitations are discussed in this paper, as well as a general framework for researchers.

In [5] they designed a system based on a skin-color modeling technique, i.e., explicit skin-color space thresholding. The skin-color range is predetermined which extract pixels (hand) from non-pixels (background). The images were fed into the model called the Convolutional Neural Network (CNN) for bracket of images.

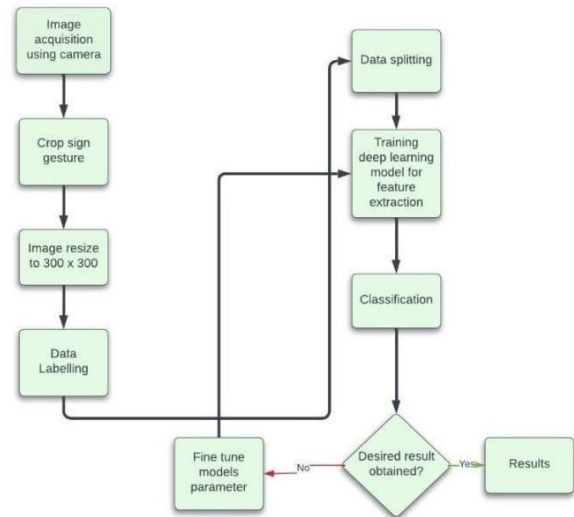
Keras was used for training of images. Handed with proper lighting condition and a invariant background, the system acquired an average testing delicacy of 93.67, of which 90.04 was attributed to ASL ABC recognition, 93.44 for number recognition and 97.52 for stationary word recognition.

PROPOSED SYSTEM

Real time sign language to textual content and audio translation, specifically 1 Reading sign gestures 2 Training the system learning model or image to textual content translation. 3 Obtaining the audio output.

A. FLOW DIAGRAM

The flowchart explains the steps occurring to accomplish the objectives of the project.



B. SYSTEM DESIGN AND RATIONALE

3.1 Dataset Preparation

The major problem in the field of hand gesture recognition for Malayalam sign language is a lack of publically available dataset, thus to overcome this problem a large dataset is collected from multiple signers under different light and background condition. For the experimental work, we have used Malayalam Sign language image dataset comprising of 56 classes with approximately 3000 instances in each class. In the dataset, each class refers to a different letter in Malayalam sign language. Each image in the dataset has dimension 224 x 224 px. Dataset contains the grayscale pixel values of the image containing the sign.

3.2 Image Acquisition

The gestures are captured to the web camera, OpenCV video stream is used to capture the entire signing duration. The frames are extracted from the stream and then processed as grayscale images.

3.3 Image Preprocessing

The captured images are scanned for hand gestures. It is a part of preprocessing before the image is fed to the model to obtain the prediction. The segment containing gestures are made more pronounced.

3.4 Hand gesture Recognition

The preprocessed images are fed to the inception v3 CNN model. The model that has already been trained generates the predicted model. The label which has highest

probability is treated to be the predicted label.

3.5 Display as Text and Speech

The model accumulates the recognized gesture to text. The recognized text are converted into corresponding audio pyttx3 library.

C. IMPLEMENTATION

In this proposed system we will focus on an efficient deep neural network architecture for computer vision, inception. Inception network was considered as a deep convolution architecture for solving image recognition and detection problem. It is basically a CNN with 27 layers deep.

In this work we have classified our Malayalam sign language letters using inceptionV3.

InceptionV3 model Architecture

InceptionV3 model has a total of 42 layers and a lower error rate than its predecessors. The inceptionv3 is just the advance and optimized version of the inceptionV1 model. In inceptionV3 model have following modification:

1. *Factorization into smaller convolution* One of the major assets of the inceptionV1 model was the generous dimension reduction. To make it even better, the larger

convolutions in the model were factorized into smaller convolution.

2. *Spatial factorization into asymmetric convolutions* Even though the larger convolutions are then factorized into smaller convolutions. A better alternative to make the model more efficient is Asymmetric convolutions. Asymmetric convolution are of the form $n \times 1$. So, replace the 3×3 convolution with a 1×3 convolution followed by a 3×1 convolution.

3. *Utility of Auxiliary classifiers*

The objective of using an auxiliary classifier is to improve the convergence of deep neural networks. The auxiliary classifier is mainly used to combat the vanishing gradient problem in very deep networks. The auxiliary classifier didn't result in any improvement in the early stages of the training. But towards the end, the network with auxiliary classifier showed higher accuracy compared to the network without auxiliary classifiers. Thus the auxiliary classifier act as a regularizer in inceptionV3 model architecture.

4. *Efficient Grid Size Reduction*

Traditionally max pooling and average pooling and average pooling were used to reduce the grid size of the feature maps. In the inceptionV3 model, in order to reduce the grid size efficient the activation dimension of the network filter is expanded. InceptionV3 model has higher efficiency. It has a deeper network compared to the inceptionV1 and V2 models, but its speed isn't compromised. It is computationally less expensive. It uses auxiliary classifier as regularizes.

CONCLUSION

In this paper, a hand gesture recognition technique is presented for vision-based recognition of sign language. For this, a deep learning based CNN model packed with compact representation is proposed.

A finger spelling sign language translator with a 95% accuracy rate is obtained. The project can be extended to other sign languages by building the corresponding dataset and training the CNN. These architectures can be explored more to minimize the error rate in real-time recognition of sign language. The experimental result show that our model can achieve good results.

REFERENCES

- [1] Sakshi Sharma, Sukhwinder Singh Expert Systems With Applications 182 (2021) 115657 ,Elsevier.
- [2] Ankit Ojha ,Ayush Pandey ,Shubham Maurya , Abhishek Thakur , Dr. Dayananda P . Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network(2020) International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 NCAIT - 2020 Conference Proceedings
- [3] Muhammad Al-Qurishi , (Member, Ieee), Thariq Khalid, And Riad Souissi , Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues [2021], IEEE Access.
- [4] Vanita Jain , Achin Jain , Abhinav Chauhan ,Srinivasu Soma Kotla , Ashish Gautam. American Sign Language recognition using Support Vector Machine and Convolutional Neural Network (2021). Bharati Vidyapeeth's Institute of Computer Applications and Management 2021.
- [5] Yanqiu Liao, Pengwen Xiong ,Weidong Min ,Weiqiong Min , And Jiahao Lu. Dynamic Sign Language Recognition Based On Video Sequence With BLSTM-3D Residual Networks(2019), IEEE Access.