JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Recognition of Sign Language in Real-Time

A Deep-Learning Based Approach

[1]Suraj Kumar B P, [2]Arathi M, [2]Harshini C M, [2]Niharika S, [2]Tejasvi Kannan

[1]Assistant Professor, [2]Student
[1]Department of Computer Science and Engineering,
[1]Sir M Visvesvaraya Institute of Technology, Bengaluru, India

Abstract : Sign Language is a form of communication used primarily by people hard of hearing or deaf. This language enables you to interact with a variety of hearing, hard-of-hearing, and deaf people and forms of difficulties caused by hearing issues. Hence, these types of gesture-based languages allow people to convey ideas and thoughts by easily overcoming these barriers that may arise.

In this study, methods of deep learning have been deployed to build a system that can detect facial expressions and signs. The system will use a camera to capture the live video of a person to analyse both signs and facial expressions. The sign language gesture recognition module will be designed to recognize individual signs and translate them into text. The facial expression recognition module will detect emotions in real-time, such as happiness, sadness, anger, surprise, fear, disgust, and neutral. The model is developed using Python, OpenCV, and deep learning frameworks such as TensorFlow and Keras. The accuracy of the system will be evaluated using real-world data, and its usability will be assessed through user testing. The main aim is to develop an effective and dependable system that can improve the standard of communication for people with speech and hearing impairments.

***IndexTerms* – Sign Language Recognition, Hand Gestures, Computer Vision, Deep Learning, Emotion Recognition, Neural Networks, LSTM.**

I. Introduction

This paper presents a mechanism that tackles the issue of communication among those who are hard of hearing by providing a machine learning-based recognition of sign language. Sign language is a non-verbal language that uses manual and non-manual indicators. Manual signs include hand and finger motions, hand posture, and gestures, among other things. Non-manual signs include head and lip motions, non-affective facial expressions, etc. These give the manual indicators important linguistic context. There is no standardization in sign language because communication varies between nations.

The National Association of the Deaf reports that there are 18 million hearing-impaired Indians. However, the Indian Journal of Otology states that it is likely to be in the neighbourhood of 63 million. Our work in this project is concentrated on offering a real-time communication system to assist deaf and mute persons in India.

Our dataset is gathered and created using a video-capturing device. First, the images are collected, then using pre-processing techniques we extract information from the frames and then proceed by using neural networks to map each frame into a matching word in our dataset. Then we continue with training the images obtained with the focus to achieve increased accuracy. The output of the model is classified into a hand gesture class from the dataset. This output is displayed as text to the user. A system to detect emotions is integrated with this to enhance the model.

II. LITERATURE SURVEY

Dhivyasri. S, Krishnaa Hari. K B, Akash. M, Sona. M, Divyapriya. S and Krishnaveni. V (2021) [1] created a system that converts gestures to text and vice-versa. This conversion involves image pre-processing, dataset collection, segmentation, feature extraction, and classification. The feature extraction process makes use of Speeded-Up Robust Feature (SURF). Machine learning techniques like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and SVM are used to compare the output with the dataset. PyAudio and the Google Speech Recognition API are used to perform the speech-to-gesture conversion. Speech is translated into text, and then the resulting text is checked against a database. After that, the suitable output is shown.

The paper by **I. Dhall, S. Vashisth and G. Aggarwal (2020)** [2] aims to demonstrate an algorithm to recognize images of a variety of hand signals and gestures, such as a thumbs-up, a closed fist, a finger count, etc. Convolutional Neural Networks with Keras and TensorFlow will be the class utilized for deep learning as visual image analysis is being employed. This study makes use of data collected with the OpenCV package and intends to enhance the accuracy rating of the current approaches for the detection of hand gestures.

This study by **M. Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, Hassan Mathkour and Mohamed Amine Mekhtiche (2020) [3]** examines a novel system suggested for the identification of dynamic gestures of hand using multiple deep learning architectures. An immensely difficult data is given as input to test the suggested system in an uncontrolled scenario. The estimate and detection of hand regions were done using the OpenPose framework. The body parts ratios theory and an efficient face detection technique were used to estimate and normalize gesture space. For learning the fine-grained characteristics of the hand shape and the coarse-grained features of the global body configuration, which is extremely effective for complex structured hand movements of sign language, two 3DCNN instances were employed separately.

Balaji Balasubramaniam, P. Diwan, R. Nadar and A. Bhatia (2019) [4] examined numerous techniques and algorithms, ranging from SVM to CNN, utilized for Facial Emotion Recognition by reading fundamental research publications. The SVM adjusts the capacity of the classification function by increasing the margin between the training patterns and the decision boundary to the maximum. When class distributions are Gaussian, Linear Discriminant analysis outperforms SVM at making predictions by calculating the likelihood that a given set of inputs belongs to each class. A set of unknown variables can be predicted using the Hidden Markov Model using a set of known variables. Convolutional Neural Network outperforms algorithms such as SVM.

The approach by **R. Pathar, A. Adivarekar, A. Mishra and A. Deshmukh (2019) [5]** aims to classify a facial image into one of the seven emotions, by building a multi-class classifier. Image acquisition is real-time using the web camera. The CNN model is deployed using the grayscale images from the FER2013 dataset. The accuracy obtained is estimated to be 90%. The effectiveness of current deep networks and shallow networks is analysed in recognition of human emotion.

III. RESEARCH OBJECTIVE

The ability to express complex and precise meaning, on par with spoken language, is the essence of sign language. That is to say, a signer is effectively capable of communicating any meaning that can be expressed in natural language. Our visual system has a complex relationship with emotion. In some aspects it is very involved: our visual cortex is activated to help us process and identify emotions. Hence, signing effectively can at most only express the meaning and not necessarily the human feelings at that moment. Sign language can be utilised to its fullest when both gestures and facial expressions are used in order to convey information. The expression of emotion on a person's face reveals a lot about their mental processes and provides a window into their inner thoughts. Real-time emotion recognition gives computers a human-like understanding of how to identify and interpret human emotions. The fact that so few people are able to develop this capability is a constraint. It is primarily limited to those who must use sign language because they cannot (easily) communicate in spoken form, i.e., primarily those who are deaf. Hence there is a need for systems that recognize the different signs and convey the ideas and thoughts of the impaired to normal people as well as classify a facial image into one of the seven emotions derived from FER2013 dataset. We are considering creating a multi-class classifier. We describe a real-time implementation of emotion and gesture identification in a web camera that produces precise results for several faces concurrently.

IV. METHODOLOGY

Our proposed system adopts two parallel procedures, one for sign language recognition and another for facial emotion recognition, as shown in Fig. 1. A sequential model with LSTM layers is trained to recognize 6 sign language gestures.

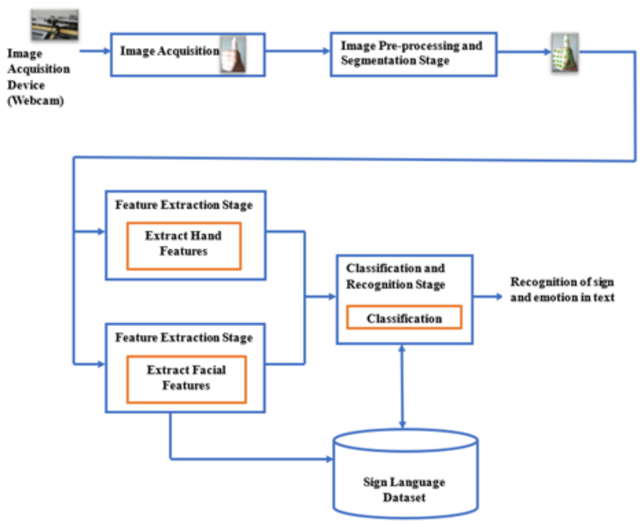

Figure 1 System architecture

A model with convolutional layers is trained separately to identify the 7 major facial emotions. The two models are then combined to provide real-time gesture and emotion recognition.
The procedure to detect hand gestures has been divided into 3 sub-modules:
1. Extracting key holistic points: We have come up with a distinct dataset, which has 30 sequences, with 30 frames each for every gesture. Collecting and extracting key holistic points is done from MediaPipe extraction. Data of different key points on hands, body, and face is collected and saved as NumPy arrays.
2. Training the LSTM DL model: A deep neural network is trained using LSTM layers to recognise the temporal components of hand gestures. The actions are predicted using a number of frames instead of a single frame.
3. Making real-time predictions using sequences: Using the previously trained model, real-time predictions of sign languages and emotions are made with OpenCV using a webcam.

The procedure to recognise facial emotions is also divided into 3 sub-modules:

1. Data Preparation: Initially we gather and prepare a dataset of labelled facial images. The FER2013 dataset is used which includes numerous people with different facial expressions. The data is pre-processed so that it becomes suitable for training by resizing the images, converting them to grayscale, and breaking down the data into training and testing sets.
2. Model Training: This typically involves defining the CNN architecture using Keras, with fully connected, pooling layers, and convolutional layers and compiling the model by specifying the loss function, optimizer, and evaluation metrics. The model is then trained using the training dataset.
3. Evaluation and Integration: The system is evaluated using the testing dataset to measure its performance and accuracy. The trained model is then integrated with the sign detection model to provide the necessary predictions.

Figure 2 Emotion recognition dataset with image from each emotion class

V. RESULTS AND CONCLUSION

The goal of our model was to create a system that can accurately understand and interpret the emotions expressed in sign language gestures. It outlines sign language recognition technology and face emotion detection that is incorporated to improve analysis and communication with the differently abled. Different algorithms' training times and levels of accuracy were compared, and the best algorithms were chosen. The system's accuracy and usefulness may be further improved, and future work in this area could foster its integration with other platforms and systems.

Because facial expression to sign language is like tone to speech, combining the two modules will result in a more thorough understanding of communication, which can be helpful for people with hearing and speech impairments.

REFERENCES

[1] Dhivyasri. S, Krishnaa Hari. K B, Akash. M, Sona. M, Divyapriya. S and Krishnaveni. V, "An Efficient Approach for Interpretation of IndianSign Language using Machine Learning," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp.130-133,doi: 10.1109/ICSPC51351.2021.9451692.

[2] I. Dhall, S. Vashisth and G. Aggarwal, "Automated Hand Gesture Recognition using a Deep Convolutional Neural Network model", 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 811-816, doi:10.1109/Confluence47617.2020.9057853.

[3] Al-Hammadi, Muneer & Muhammad, Ghulam & Abdul, Wadood & Alsulaiman, Mansour & Bencherif, Mohamed & Alrayes, Tareq & Mathkour, Hassan & Mekhtiche, Mohamed, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation", 2020 IEEE Access. 8. 192527-192542.10.1109/ACCESS.2020.3032140.

[4] B. Balasubramanian, P. Diwan, R. Nadar and A. Bhatia, "Analysis of Facial Emotion Recognition", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 945-949, doi: 10.1109/ICOEI.2019.8862731.

[5] R. Pathar, A. Adivarekar, A. Mishra and A. Deshmukh, "Human Emotion Recognition using Convolutional Neural Network in Real Time," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-7, doi: 10.1109/ICIICT1.2019.8741491.