e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

SIGN LANGUAGE TRANSLATION USING CNN SURVEY

Tanmay Dabhade*1, Swati Ghawate*2, Aishwarya Diwane*3,
Cedrick Andrade*4, Pranav Chavan*5
*1,2,3,4,5Computer Department Savitribai Phule Pune University Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS30981

ABSTRACT

The project aims to create a machine-learning model that can classify the numerous hand gestures used for sign language fingerspelling. Classification machine learning algorithms are taught using a set of image data in this user-independent model, and testing is done on a completely diverse bundle of information Depth photos is employed for the image dataset, and they produced better results than some due to the quicker pre-processing, of the prior kinds of literature[4]. The datasets are subjected to machine learning techniques, such as Convolutional Neural Networks (CNN). The CNN model is pre-trained using the Imagenet dataset in an effort to improve its accuracy. But only a short dataset was employed for pre-training, resulting in a 15% accuracy.

Keywords: Sign Language Recognition, SVM, KNN, Logistic Regression, Convolutional Neural Networks.

I. INTRODUCTION

Communicating with those who have hearing loss is really difficult. Since Deaf and Mute people use hand gestures to communicate, normal people have difficulty understanding their language from the signals they make. Systems that can identify various indications and provide information to common people are thus necessary [1], [2].

The majority of individuals have trouble understanding ISL gestures. As a result, there is now a communication barrier between individuals who comprehend ISL and those who do not. When necessary, it can be difficult to locate an interpreter who can translate these signals. A prospective solution that would translate hand positions and gestures from ISL in real time was put into place to help with communication. It includes a server and an Android smartphone camera for capturing hand positions and motions [3], [4].

We investigate different machine learning techniques like Support Vector Machines (SVM), Logistic Regression, K-nearest neighbors (KNN) and a neural network technique Convolution Neural Networks (CNN) for detection of sign language.

Standard hand-based gestures are part of the Indian Sign Language system, which is used by people with speech impairments to communicate.

Due to their intricacy and breadth, these gestures are difficult for many people to understand, which hinders communication between those with and without speech impairments. There is a lot of active practical research being done in the field of computer vision as a result of the current boom in deep learning. Despite this, there has only been a small amount of inadequate research done on gesture recognition in ISL [1]. Our paper focuses on creating a baseline for ISL gesture identification and developing a model to help the communication of the speech-impaired with the intention of improving the same.

II. KEY CONCEPTS

A. Image processing

Image processing is a technique for applying certain operations to an image to produce an improved image or to extract some relevant information from it. It is a form of signal processing in which a picture serves as the input, and the output could be another image or characteristics or features related to that image.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

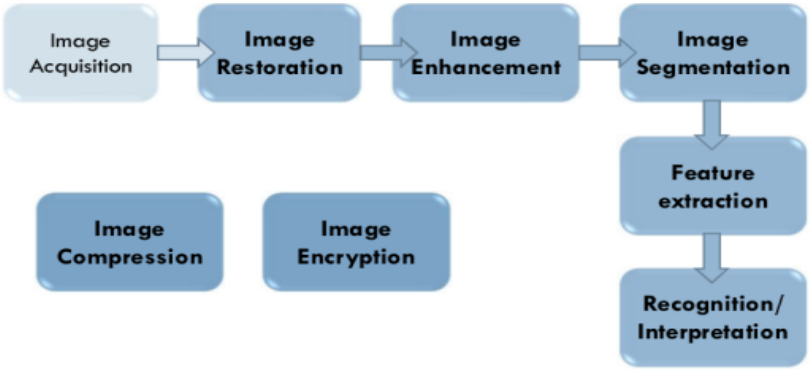

Fig: Image processing

B. Video processing:

Video transmissions from security cameras are encoded or encrypted during video processing. The processing enables the display of many cameras on a single monitor or the cycling of several camera views on a monitor. A cross-point matrix and multiplexer have generally been employed to do this, both of which are relics of analogue video systems.

The older cross-point matrix's primary function is to interchange camera feeds between several outputs. It "binds" any input to any output, or to several outputs, using the feeds or inputs from the video security cameras.

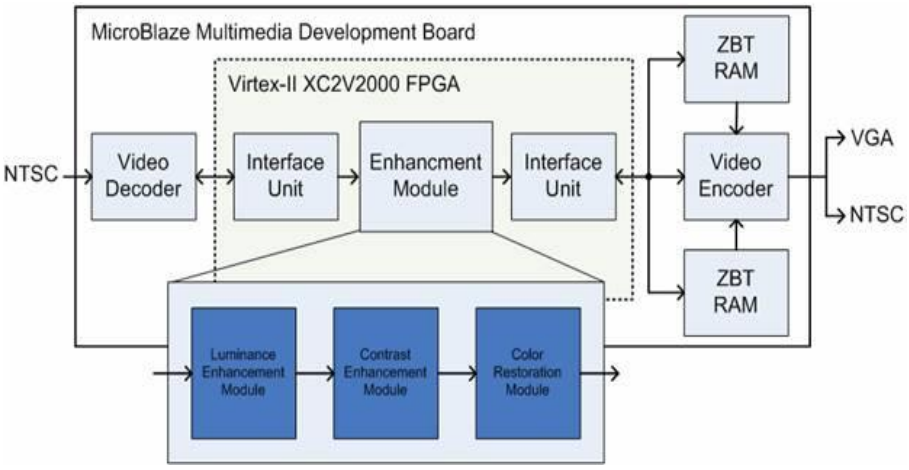

Fig: Video processing

The functioning of older video multiplexers can be summed up as taking several video streams and combining them into one video feed. This enables the viewing of several cameras on a single screen.

III. RELATED WORK

Speech-impaired individuals in India use the typical hand-based motions used in the Indian Sign Language system to communicate. Due of their intricacy and breadth, these gestures are difficult for many people to understand, which hinders communication between those with and without speech impairments. There is a lot of active practical research being done in the field of computer vision as a result of the current boom in deep learning. Despite this, there has only been a small amount of inadequate research done on gesture recognition in ISL. Our paper focuses on creating a baseline for ISL gesture identification and developing a model to help the communication of the speech-impaired with the intention of improving the same [4].

Since there are deaf people everywhere, it is imperative that local level sign language recognition (SLR) techniques be developed. Based on a thorough analysis of machine/deep learning methods and approaches for automated sign language recognition that was published between 2014 and 2021, we came to the conclusion that the present systems need conceptual classification in order to properly understand all of the available data. As a result, we focused on components that practically all sign language detection techniques share. In this

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

essay, we compare their relative merits and shortcomings and offer a broad framework for scholars. This study also demonstrates the importance of input modalities in this area; it seems that recognition based on a combination of data sources, including vision-based and sensor-based channels, is superior to a unimodal approach [5].

While there are many different gestures, sign languages have the most structured sets. Each gesture in sign language already has a specific meaning, and strong context and grammar rules can be used to make identification manageable. The majority of deaf people in the United States prefer to communicate using American Sign Language (ASL). ASL employs about 6,000 gestures to communicate simple words and finger spelling for more difficult words or proper nouns. The bulk of signing, however, uses complete words, which enables signed discussions to move along at a rate similar to spoken conversations [9].

IV. METHODOLOGY

Modern technology is sophisticated and capable of helping the deaf and hard of hearing. We propose a model that can comprehend sign language and uses the Convolutional network The predicted alphabet will be shown in real-time when sign language is detected using a smartphone camera. The model can also be made available for a variety of sign languages

To build this model, we will be using the Keras and Tensorflow frameworks. We are going to create the dataset for this by retrieving the images or frames from the film using the OpenCV library.

Data collection is the initial stage of the proposed system.

In numerous investigations, sensors or cameras have been used to record hand movements.

WHAT IS SIGN LANGUAGE?

Sign language is a visual language. It mainly consists of 3 major components:

Fingerspelling: Spell out words character by character, and word level association which involves hand gestures that convey the word meaning. The static Image Dataset is used for this purpose.

World-level sign vocabulary: The entire gesture of words or alphabets is recognized through video classification. (Dynamic Input / Video Classification)

Non-manual features: Facial expressions, tongue, mouth, body positions.

Data Collection & Pre-Processing:

We will be using 2 datasets and then compare the results.

Dataset 1:

MNIST Dataset : 28×28 pixels images(24 alphabets: J and Z deleted as they include gesture movements: (Dataset) [Training: 27,455 , Testing: 7172]

Dataset 2:

Image Dataset: 200×200 pixels images: 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE, and NOTHING. Dataset (J and Z were converted to static gestures by converting only their last frame).

Applications for computer vision range from useful ones in industry to useful ones in society. It has also been used in numerous programmes to assist persons who are physically handicapped. Computer vision can produce English alphabets for deaf-mute people based on sign language symbols.

In this tutorial, we'll use a convolutional neural network to categorise the signs used in sign language (CNN). The associated alphabet of a sign language symbol will be predicted following successful CNN model training. We will use the non-normalized and normalised confusion matrices to assess the classification performance of our model. Finally, we will determine the CNN model's classification accuracy score for this assignment.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

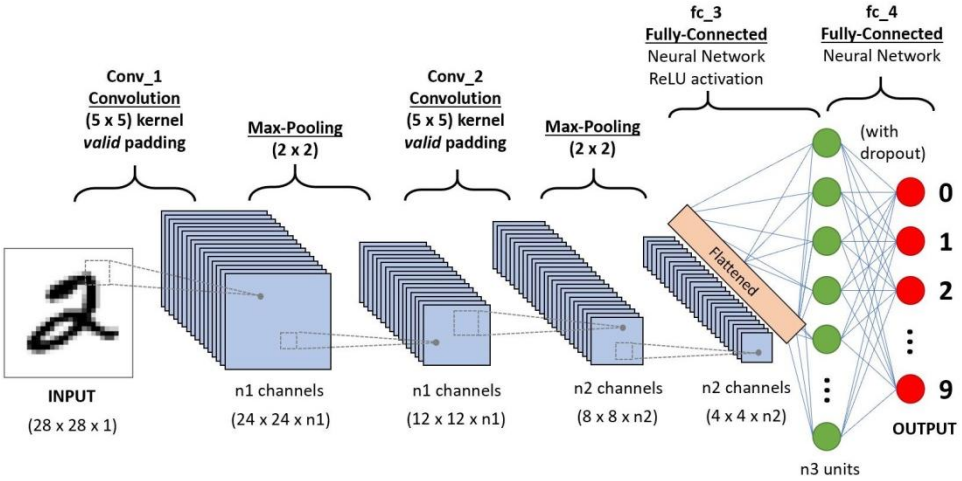

Fig: CNN Architecture

V. CONCLUSION

There have been significant developments in computer vision, machine learning, and artificial intelligence. We now apply their techniques much more successfully and have a more expanded perspective on the world.

Numerous studies have been conducted on the recognition of sign gestures using various methodologies such as ANN, LSTM, and 3D.

ACKNOWLEDGEMENT

We would like to extend our gratitude to Project staff at the Sinhgad Academy of Engineering, Department of Computer Engineering for their support and help without which this work could not have been accomplished.

Additionally, we would like to give Mrs. Swati Ghawate, our project manager, our profound gratitude for leading us through the process.

VI. REFERENCES

[1] Q. Wu, Y. Liu, Q. Li, S. Jin and F. Li, "The application of deep learning in computer vision," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 6522-6527.

[2] D. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition, vol. 13, no. 2, pp. 111122, 1981.

[3] D. G. Lowe,Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, vol. 13, no. 2, pp. 111122, 1981.

[4] Neel Kamal Bhagat, Y. Vishnusai and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning", 2019 Digital Image Computing: Techniques and Applications (DICTA).

[5] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep learning for sign language recognition: Current techniques benchmarks and open issues", IEEE Access, vol. 9, pp. 126917-126951, 2021.

[6] M. J. Cheok, Z. Omar, and M. H. Jaward, ''A review of hand gesture and sign language recognition techniques, ''Int. J. Mach. Learn. Cybern., vol. 10, no. 1, pp. 131–153, Jan. 2019.

[7] Q. Xiao, Y. Zhao, and W. Huan, ''Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and con volutional neural network, ''Multimedia Tools Appl., vol. 78, no. 11, pp. 15335–15352, Jun. 2019.

[8] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, and D. A. Kumar, ''3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2–stream CNN, ''Neurocomputing, vol. 372, pp. 40–54, Jan. 2020.

[9] Thad Starner, Joshua Weaver and Alex Pentland, "Real-time american sign language recognition using desk and wearable computer based video", IEEE Transactions on pattern analysis and machine intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.