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

A Review on Scrutinizing Sign Language Interpreters

[1]Aryan Samsukha, [2]Siddhanth Suresh, [3]Sanjana Natraj

[1]Student, [2]Student, [3]Student

[1]BCA-DA,

[1] *JAIN(Deemed-to-be) University, Bengaluru, Karnataka*

Abstract : ***Inability to speak is a true disability and there are various methods of communicating and among one such method is sign language . Understanding sign language is one such important task and it may involves a interpreter ,which may costs more . To ease the use of interpreter we created a software which is free of use and which can be modifiable for every interactive user. This project's purpose is to create a vision-based system that can recognise sign language motions in video sequences. A vision-based approach was chosen because it provides for a more straightforward and intuitive flow of data between a human and a computer. The programme has a 93 percent accuracy rate in terms of quality.***

KEYWORDS: ASL (American Sign Language), ISL (Indian Sign Language), CNN (Convolutional Neural Network), RNN (Recursive Neural Network), LSA (Argentinian Sign Language).

I. INTRODUCTION

Many people around the world are born with disabilities like dumb, deaf and mute. They each have their own communication styles. There is one common technique of communication among a variety of communication modalities, and that is sign language.

A technique of communication based on visual motions and signs is known as sign language. Sign language is divided into several groups, including ISL (Indian Sign Language), ASL (American Sign Language), and others (American Sign Language). A person should understand the language to convey to the person. To overcome this issue a software is created to understand sign languages on behalf of humans. This current topic is being promoted by the government and foster the inclusion of hearing-impaired people and providing them with the new environment.

To see a person who is unable to speak can stand and make gestures, and the system can turn the movements into words. Earlier the projects were only focusing on the static images and our works include video recognition system with only one hand detection.

Human-computer interaction, computer vision, movement analysis, automated learning, and pattern recognition are just a few of the fields where signal complexity inspires research [1].

Although there are many possible methods to use amongst, CNN is used as it requires feature extraction and it is much powerful and are prevalent when using images and video processing [3].

Sign language is a visual language that consists of three main elements:

Fingerspelling	**Word level sign vocabulary**	**Nonmanual features**
Letter by letter , words are spelled	Hand gestures are made to perform word level signs .	Facial expressions and mouth and body positions.

LITERATURE SURVEY

The majority of existing SLI can only recognise single alphabets from hand gestures, which is the root of the problem, which is referred to as finger spelling. Finger spelling is not a feasible option because it reduces SLI's efficiency owing to frequent delays [15].

The research may not include all of the signs proposed in the American sign language (ASL) but a few of them have been realized [2]. Our Software tried to remove the limitations possessed and tried to make faster. Instead of using a wearable gadget, this proposed system will use human motion and gesture recognition for carrying out the conversation. This system removes the use of gloves while describing the use of the software [4]. A high-level programming language like Python is used to build a GUI for interaction with the user [12].

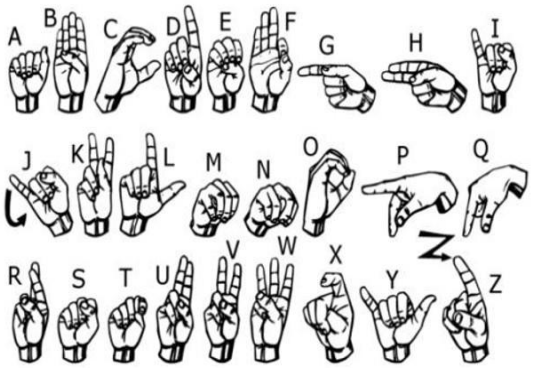

Figure 1: Sign-conventions

Now the dataset used for building the model has been named according to the proper interpretation of the user. A distinct name is assigned to each of the 26 possible English alphabets in the dataset [6].

To begin, videos of people communicating in ASL are converted into frames using the ffmpeg API, with 1 frame per second frames extracted. Then each frame is stored with a distinct frame IDentification. Furthermore, the collected frames are pre-processed to convert them to greyscale images using pre-processing techniques such as Skin Filtering. In order to detect hand gestures, the input frames are subjected to a skin filtering approach. This is done so that the necessary hand motion can be extracted from the background.
Skin Filtering is a method of distinguishing between skin-colored and non-skin-colored areas. The cánny edge detection álgorithm is used to filter the skin. [6].

The next step is to find the feature description and then to detect the feature point. The Scale Invariant Feature Tránsform (SIFT) features of diverse American Sign Language alphabets are then used to build the Convolutional Neural Network (CNN) classification model, which has 26 classes [5].

SVM, which uses supervised learning for sign recognition and uses the test data as the input, is used to evaluate the classification model [15]. The best gestures are identified by analysing the HMMs that have been trained. Finally, depending on the motion provided as input, The output is available in text or speech format. While the user can add their own data sets and train them for extra flexibility.
There are many publications and research papers in the recent areas , describing the system, but none of them prove to make a more efficient and reliable working product which is useful for the community.

PROPOSED SYSTEM

INITIALIZING

The methods requires a camera and a natural interaction between humans and computers using the software. This is a tough problem to tackle because, in order to achieve real-time performance, these systems must be background insensitive, illumination insensitive, and person and camera agnostic.

Image Acquisition

The first part is to create a database of handshapes for the American Sign Language (ASL). The handshapes are used with different angles and postures with the same action. This may create accuracy in the software. A region that is made to primarily focus on the visual patch that surrounds a hand gesture in the application of hand gesture recognition [10].
The client is forced to make gestures in fore of the camera only by making use of a webcam, which are then sent to the program for processing, which are then sent to the program for processing. It is expected that the camera will be fixed. Real-time limitations are set in order to ensure that the processing system is carefully designed. The background should be checked properly for any disturbances and camera must be standard pixels size. Now these images has been captured and are ready for the pre-processing.

Pre-Processing :

Removing unnecessary factors from the image captured like background disturbances, using the suppressing the background and focusing on the human areas First, the RGB image was transformed to HSV colour system, which is more closely related to human colour perception. Channel H's skin is defined by values ranging from 0 to 50 [7]. So any part of the image or the components not required for the processing are being separated out.

Now the problem faced here is continuous video sequences and it is solved using key frame extraction method by splitting each sign into sequence of signs.

Feature Extraction :

The second stage is to measure the contour of the gesture based on the shape generated by the hand.

The contour determined from the image is matched to the database's appropriate sign.

Using thresholding, create a binary version of the processed image. The imáge segmentation method can be used to perform the binary conversion [9]. The pixels in the background can be transformed to black colour pixels, while pixels in our region of interest can be turned to white colour pixels.

4.Clássificátion :

Images, their matching IDs, and the sign's written meanings are all stored in the database. Our code then compares the image to the associated Id. After the text-to-sound converter reads from a text file, the accompanying sound is created.

The user will have the option of adding additional signs to the dátábáse in order to make the proposed system dynamic. To make things even easier, sound has been added to the database, as well as the ability to insert meáning for a new sign as a string of text.

Testing :

This process involves checking the software by evaluating the system with some pre-defined requirements mentioned in the documentation of the project. Some of the factors to be assessed are the accuracy of gesture recognition, Database dynamicity, translation from text to sign, insertion of customised signs, and ease of use are all features that can be found in the conversion from sign to text and speech.

Convolutional Neural Network

It is made up of a network of neurons, which are learning units. These neurons learn how to convert input signals (such as a cat image) into output signals (for example, the label "cat"). In CNNs, repetitive blocks of neurons are used all over the place [8].

In terms of pictures, these blocks of neurons can be thought of as 2D convolutional kernels applied repeatedly over each patch of the image. They can be compared to 1D convolutional kernels applied to a voice temporal window [11].

CNN IS DIVIDED INTO FOUR STEPS:

- CONVOLUTION: Convolution, also known as Translation invariant, is a process in which the network attempts to classify the incoming signal by referencing what it has previously learned. The output signal is then passed on to the next layer [14].
- Subsampling: The convolution layer's inputs can be "smoothed" to lessen the filters' sensitivity to noise and fluctuations. Sub-sampling is the term for the smoothing procedure. This can include shrinking the image or diminishing the colour contrast in the RGB channels (red, green, and blue).
- POOLING: Its goal is to reduce the number of parameters and calculations in the network by gradually shrinking the spatial dimension of the rep-resentation.
- ACTIVATION: More neurons would be activated if the output signals are strongly related to it, enabling signals to be propagated more efficiently for identification [13].
- FULLY CONNECTED: Every neuron in the following layer is related to the neurons in the previous layer.

CONCLUSION

This project is a minor start toward assisting physically challenged persons, and there is much more that can be done to improve the product's sophistication, usability, and efficiency. Additional languages can be supported with more memory and a strong microprocessor. This project can be tweaked to make it mobile phone friendly. Using a more capable trans-receiver module, we can expand the product range. A vision-based system capable of interpreting individual hand movements from the American Sign Language was shown in this research (ASL).

ACKNOWLEDGMENT

We would like to thank our mentor **Dr. Mahesh V.** (Assistant Professor, Department of Computer Applications) for giving us the opportunity to undertake the project. We thank them for their immense guidance, and appreciate their timely engagement.

REFERENCES

[1] Ronchetti, Franco, Facundo Quiroga, César Armando Estrebou, and Laura Cristina Lanzarini. "Handshape recognition for Argentinian sign language using probsom." Journal of Computer Science &Technology 16 (2016).

[2]. Singha, Joyeeta, and Karen Das. "Automatic Indian Sign Language Recognition for Continuous Video Sequence." ADBU Journal of Engineering Technology 2, no. 1 (2015).

[3]. Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. "Learning longterm dependencies with gradient descent is difficult." IEEE transactions on neural networks 5, no. 2 (1994): 157166.

[4]. .N.Hema,Ms. P. Thamarai ,Dr. T.V.U. Kiran Kumar,"HandiCom: Handheld Deaf and Dumb Communication Device based on Gesture to Voice and Speech to Image/Word Translation with SMS Sending and Language Teaching Ability, "IJESRT,ISSN: 2277 - 9655,Department of Electronics and Communication Engineering, Tamil Nadu, India

[5] . Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25

[6] Prateek S.G.;Jagadeesh J.;Siddarth R.;Smitha Y.;P. G. Sunitha Hiremath;Neha Tarannum Pendari 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)

[7] A. Albiol, L. Torres, and E. J. Delp. "Optimum color spaces for skin detection." In proceedings of the 2001 international conference on image processing, volume 1, vol. 1, pp. 122-124 ,2001.

[8] Aloimonos, Y. (1991). Computer Vision. aaas, 253(5025), 1249-1254.

[9] A Bala, "An improved watershed image segmentation technique using MATLAB", *International Journal of Scientific & Engineering Research*, vol. 3, no. 6, pp. 1-4, 2012

[10] Zhang, Chenyang, Xiaodong Yang, and YingLi Tian. "Histogram of 3D facets: A characteristic descriptor for hand gesture recognition." In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 18. IEEE, 2013.

[11] Cooper, Helen, Brian Holt, and Richard Bowden. "Sign language recognition." In Visual Analysis of Humans, pp. 539562. Springer London, 2011.

[12]. Salma A. Essam El-Din;Mohamed A. Abd El-Ghany 2020 2nd 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES) Year: 2020 | Conference Paper Publisher: IEEE

[13] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Deep learning*, *521*, 436-444.

[14] J. Singha and K. Das, "Hand Gesture Recognition Based on Karhunen-Loeve Transform", Mobile and Embedded Technology International Conference (MECON), January 17-18, 2013, India, pp. 365-371.

[15] Rewari, H., Dixit, V., & Batra, D. (2018). *Automated Sign Language Interpreter*.