A Translator for Indian Sign Language to Text and Speech

Article · June 2020

DOI: 10.22214/ijraset.2020.6267

Aaditya Chintaman Punekar
Queen Mary, University of London

iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6267

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

A Translator for Indian Sign Language to Text and Speech

Aaditya. C. Punekar[1], Saranya. A. Nambiar[2], Raunak Singh Laliya[3], Prof. Saumya. R. Salian[4]
[1, 2, 3, 4]Department of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Abstract: Verbal Communication is the only way using which people have interacted with each other over the years but the case stands different for the disabled. The barrier created between the impaired and the normal people is one of the setbacks of the society. For the impaired people (deaf & mute), sign language is the only way to communicate. In order to help the deaf and mute communicate efficiently with the normal people, an effective solution has been devised. Our aim is to design a system which analyses and recognizes various alphabets from a database of sign images. In order to accomplish this, the application uses various techniques of Image Processing such as segmentation & feature extraction. We use the machine learning technique, Convolutional Neural Network for detection of sign language. We convert the image by cropping the background and keeping only gesture, after that we convert the gesture into black & white scale in png format into 55*60 resolution. This system will help to eradicate the barrier between the deaf-mute & normal people. This system will standardize the Indian Sign Language in India. It will also improve the quality of teaching and learning in deaf and mute institutes. Just as Hindi is recognized as the standard language for conversation throughout India, ISL will be recognized as the standard sign language throughout India. The main aim of this work is serving the mankind that is achieved by providing better teaching and better learning.
Keywords: Indian Sign Language Recognition, Translation, Feature Extraction, Convolutional Neural Network, Neural Network

I. INTRODUCTION

Sign language uses manual and visual mode to convey what the person thinks, feels and experiences. There are three major sign languages in the world: American Sign Language (ASL), Pidgin Signed English (PSE) and Signed exact English (SEE). All the three standardized sign languages are in English. For the local citizens of India, a foreign language like English would never become a major sign language because people normally learn to speak in their mother tongue which is not English. On the other hand, there are a variety of sign languages present throughout India. The different parts of India have little difference in signing but the grammar remains the same throughout the country. Hence, a sign language that is standardized and can be used by anyone who is deaf and mute and understood by the normal people is mandatory.

Therefore, just like Hindi is the national language and is known by majority of the citizens, which makes communication efficient, Indian Sign Language has to be standardized. There are many institutes that teach different sign language to the teachers being trained leading to different sign languages. The crux of the problem lies in little to no knowledge about this. In a diverse country like India, unifying or standardizing a sign language is complex. The solution for this must be generic as well as adaptive taking into regard the Indian diversity. Use of an external hardware is not required in our approach.

II. METHODOLOGY

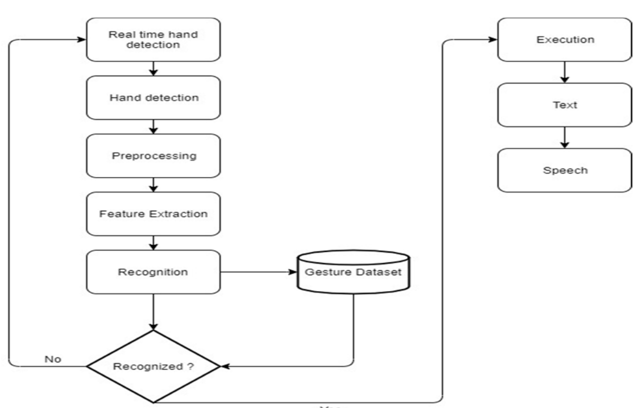


Fig. 1 Overview of the proposed system

The below figure shows the system architecture of the project.

A. Image acquisition

The initial or the first step of this system is vision based i.e. to acquire images at runtime via the camera. Then, these images will be stored in a directory, in which all the images will be stored and trained by the user and the saved & trained images will be used to compare the recently captured image to the previous images of a specific letter.

B. Feature Extraction

The palm is extracted from the images via image segmentation. The purpose of Image Processing in our project is Image Retrieval i.e. seeking for the image of interest and Image Recognition i.e. distinguishing the objects of image.

The images are first converted to grayscale from coloured. The background data being unnecessary is therefore removed as the bit size also reduces when it comes to saving the image as data which can be further trained. This is done with the help of algorithms like Gaussian Blur, Median Blur and Contour.

C. Gesture Recognition

Once the above steps are completed, the application will then match the recognized character with the already saved and trained data in our database. For the database, the user can either train one and keep it in the same application or import it via cloud example Google Colab. The Gestures are checked once the thresh is set by the user. Upon recognition, the program stands executed. Thus, displaying the text of the input sign. If the sign is not recognized, the program will be executed once again right from the start.

D. Text to Speech

As and when the character is recognized, the output will then be further converted from text to speech. Here we use pyttsx3 which is a text to speech conversion library in Python. The output text is converted to speech by pyttsx3. It works offline, unlike alternative libraries, and is compatible with both Python 2 and Python 3. We can see and hear the sign language translation in our system at that same time and that makes it convenient to use.

III. ALGORITHM USED

Convolutional Neural Networks is one of the main methods to do images recognition and image classification. CNN image classifications take an input image, process it and classify it under certain categories. Computer sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d (h = Height, w = Width, d = Dimension). E.g., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.

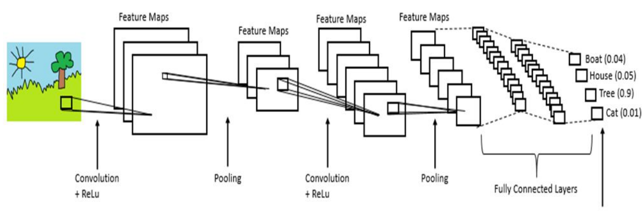


Fig. 2 CNN Architecture

A. Architecture of CNN

The main building block of CNN is the convolutional layer. Convolution is a mathematical operation to merge two sets of information. In our case the convolution is applied on the input data using a convolution filter to produce a feature map. There are a lot of terms being used so let's visualize them one by one. On the left side is the input to the convolution layer, for example the input image. On the right is the convolution filter, also called the kernel, we will use these terms interchangeably. This is called a 3x3 convolution due to the shape of the filter.

We perform the convolution operation by sliding this filter over the input. At every location, we do element-wise matrix multiplication and sum the result. This sum goes into the feature map. The area where the convolution operation takes place is called the receptive field. Due to the size of the filter the receptive field is also 3x3.

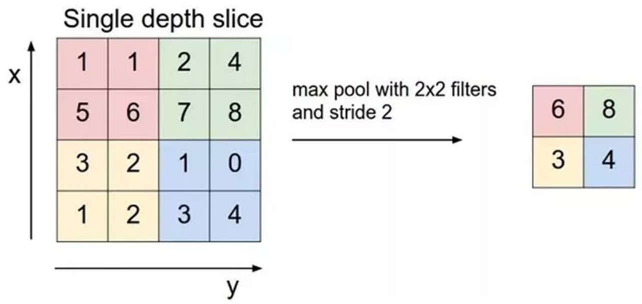


Fig. 3 Convolution Operation

The figure above shows the entire convolution operation. Here the filter is at the top left, the output of the convolution operation "4" is shown in the resulting feature map. We then slide the filter to the right and perform the same operation, adding that result to the feature map as well.

After a convolution operation we usually perform pooling to reduce the dimensionality. This enables us to reduce the number of parameters, which both shortens the training time and combats overfitting. Pooling layers down sample each feature map independently, reducing the height and width, keeping the depth intact. The most common type of pooling is *max pooling* which just takes the max value in the pooling window.

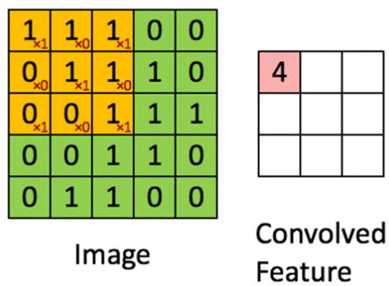


Fig. 4 Max Pooling

B. Gaussian Blur

Gaussian Blur is blurring of image by a Gaussian function. It is widely used to reduce image noise and detail. Gaussian is also used as a pre-processing stage in computer vision algorithms to enhance image structures at different scales.

The Gaussian blur is a type of image-blurring filter that uses a Gaussian function (which also expresses normal distribution in statistics) for calculating the transformation to apply to each pixel in the image. The formula of a Gaussian function in one dimension is

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

In two dimensions, it is the product of two such Gaussian functions, one in each dimension:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard of the Gaussian distribution. When applied in two dimensions, this formula produces a surface whose contours are concentric with a Gaussian distribution from the centre point. Values from this distribution are used to build a convolution matrix which is applied to the original image. This convolution process is illustrated visually in the figure on the right. Each pixel's new value is set to a weighted average of that pixel's neighbourhood. The original pixel's value receives the heaviest weight (having the highest Gaussian value) and neighbouring pixels receive smaller weights as their distance to the original pixel increases. This results in a blur that preserves boundaries and edges better than other, more uniform blurring filters.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

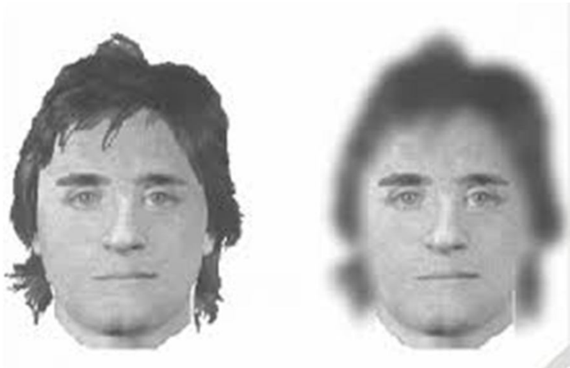


Fig. 5.1 Gaussian Blur

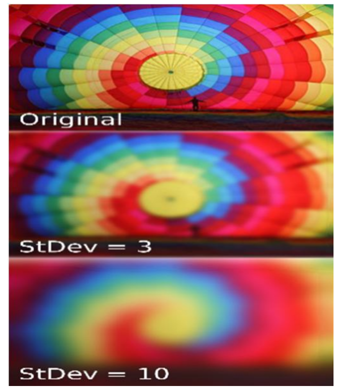


Fig. 5.2 Gaussian Blur

C. Median Filter

The Median Filter is a non-linear digital filtering technique, which is used to remove noise from an image or signal. Such noise reduction is a typical pre-processing step to improve the results of later processing for example, edge detection on an image. Median filtering is very widely used in digital image processing.

To demonstrate, using a window size of three with one entry immediately preceding and following each entry, a median filter will be applied to the following simple 1D signal:

$x = (2, 3, 80, 6)$.

So, the median filtered output signal y will be:

$y_1 = \text{med}(2, 3, 80) = 3$,

$y_2 = \text{med}(3, 80, 6) = \text{med}(3, 6, 80) = 6$,

$y_3 = \text{med}(80, 6, 2) = \text{med}(2, 6, 80) = 6$,

$y_4 = \text{med}(6, 2, 3) = \text{med}(2, 3, 6) = 3$,

i.e. $y = (3, 6, 6, 3)$.

Median filtering is one kind of smoothing technique, as is linear Gaussian filtering. All smoothing techniques are effective at removing noise in smooth patches or smooth regions of a signal, but adversely affect edges. Often though, at the same time as reducing the noise in a signal, it is important to preserve the edges. Edges are of critical importance to the visual appearance of images, for example. For small to moderate levels of Gaussian noise, the median filter is demonstrably better than Gaussian Blur at removing noise whilst preserving edges for a given, fixed window size. However, its performance is not that much better than Gaussian blur for high levels of noise, whereas, for speckle noise and salt & pepper noise (impulsive noise), it is particularly effective. Because of this, median filtering is very widely used in digital image processing.

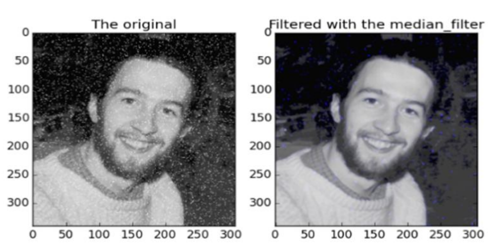


Fig. 6.1 Median Filter

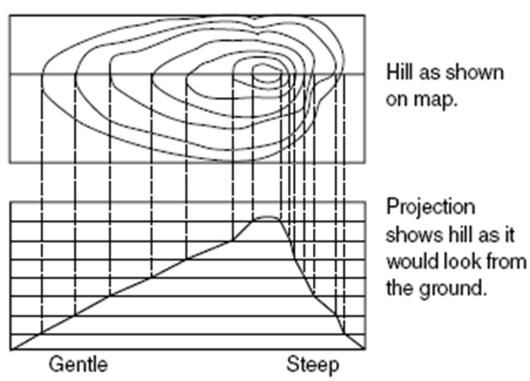


Fig. 6.2 Median Filter

D. Contour

Contour tracing is one of many preprocessing techniques performed on digital images in order to extract information about their general shape. Once the contour of a given pattern is extracted, it's different characteristics will be examined and used as features which will later on be used in pattern classification. Therefore, correct extraction of the contour will produce more accurate features which will increase the chances of correctly classifying a given pattern.

Well, the contour pixels are generally a small subset of the total number of pixels representing a pattern. Therefore, the amount of computation is greatly reduced when we run feature extracting algorithms on the contour instead of on the whole pattern. Since the contour shares a lot of features with the original pattern, the feature extraction process becomes much more efficient when performed on the contour rather on the original pattern. In conclusion, contour tracing is often a major contributor to the efficiency of the feature extraction process -an essential process in the field of pattern recognition.

Fig. 7 Contour

IV. RESULTS

The system needs no external devices, just a laptop and one can give it a go. The user can directly run the code and the system will then check the thresh which means it will check for the background, which helps in better gesture recognition. Firstly, we have to set thresh of out gesture after that we check the background is discarded in the thresh or not, if not we set the hand gesture again excluding the background or changing it. The background in the image is discarded by an image processing algorithm Gaussian Blur, Median blur and Contour and feature is extracted and pre-processing of the image is done. Once the thresh is set, the system will then check the Gesture from the database where the user has already created and trained various gestures. The gestures here are trained using Google Colab and only the trained gestures are able to show in the results. The Output file generated in the form, H5 and extension and is connected at display gesture program. After Gesture Recognition is done, the system will display the result in text as well as voice format with the accuracy of 88% with words and 96% with alphabets and numbers.

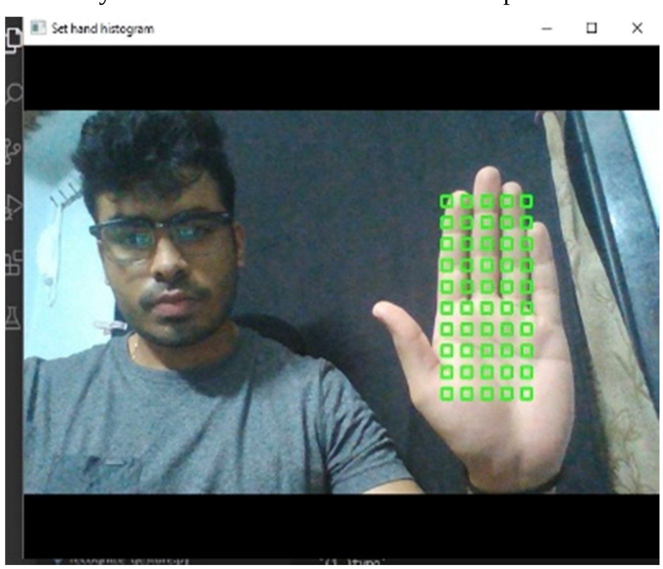


Fig. 8 Setting hand in the box

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Fig. 9 Translating Sign to Text

Fig. 10 Detection of the hand eliminating the background

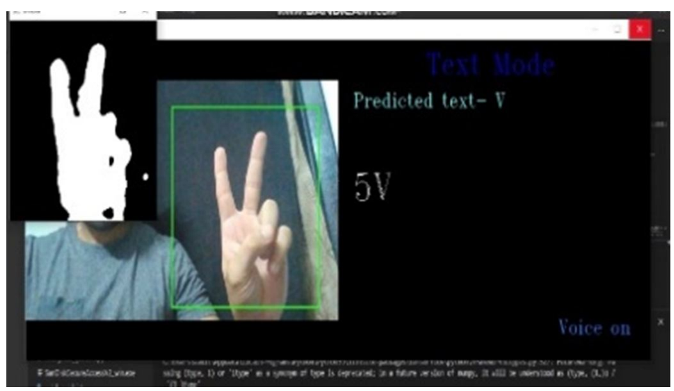


Fig. 11 Translating Sign to Text

V. CONCLUSION

The paper tries to minimize the communication barrier between people who can speak and the ones who are impaired (deaf and mute). This research if properly used and implemented can help in equalizing the number of normal and impaired students in a normal school. The communication will be far better than it used to be among students and adults. People can learn sign language from the proposed system and help the impaired people when in need. This project will help in standardizing the sign language for the deaf and mute. Now they can learn this language to communicate with each other if they belong to different regions and have been taught their regional sign language. In order to improve gesture identification, the training samples can be increased with the optimal features given as input condition for finger detection. Text to Speech conversion can be improved by already saving some common texts into the database.

REFERENCES

[1] Viraj Shinde, Tushar Bacchav, Jitendra Pawar, Mangesh Sanap - Hand Gesture Recognition System Using Camera, International Journal of Engineering Research & Technology (IJERT) Vol.3 Issue1, January-2014 IJERTISSN:2278-0181

[2] Dave Rupit, Patel Jwoli - Hand Gesture Recognition for Computer Vision Technology, International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 1, January - 2014 IJERT ISSN: 2278-0181

[3] Mayur V. Gore, Dr. Anil R. Karwankar - Human Computer Interaction using Hand Gesture Recognition, International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 4, April - 2014 IJERT ISSN: 2278-0181

[4] Arjunlal, Minu Lalitha Madhavu - A Survey on Hand Gesture Recognition and Hand Tracking, International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-2, Issue-1, January 2016 ISSN: 2395-3470

[5] Sachin Bhat, Amruthesh M, Ashik, Chidanand Das, Sujith - Translating Indian Sign Language to text and voice messages using flex sensors, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 5, May 2015

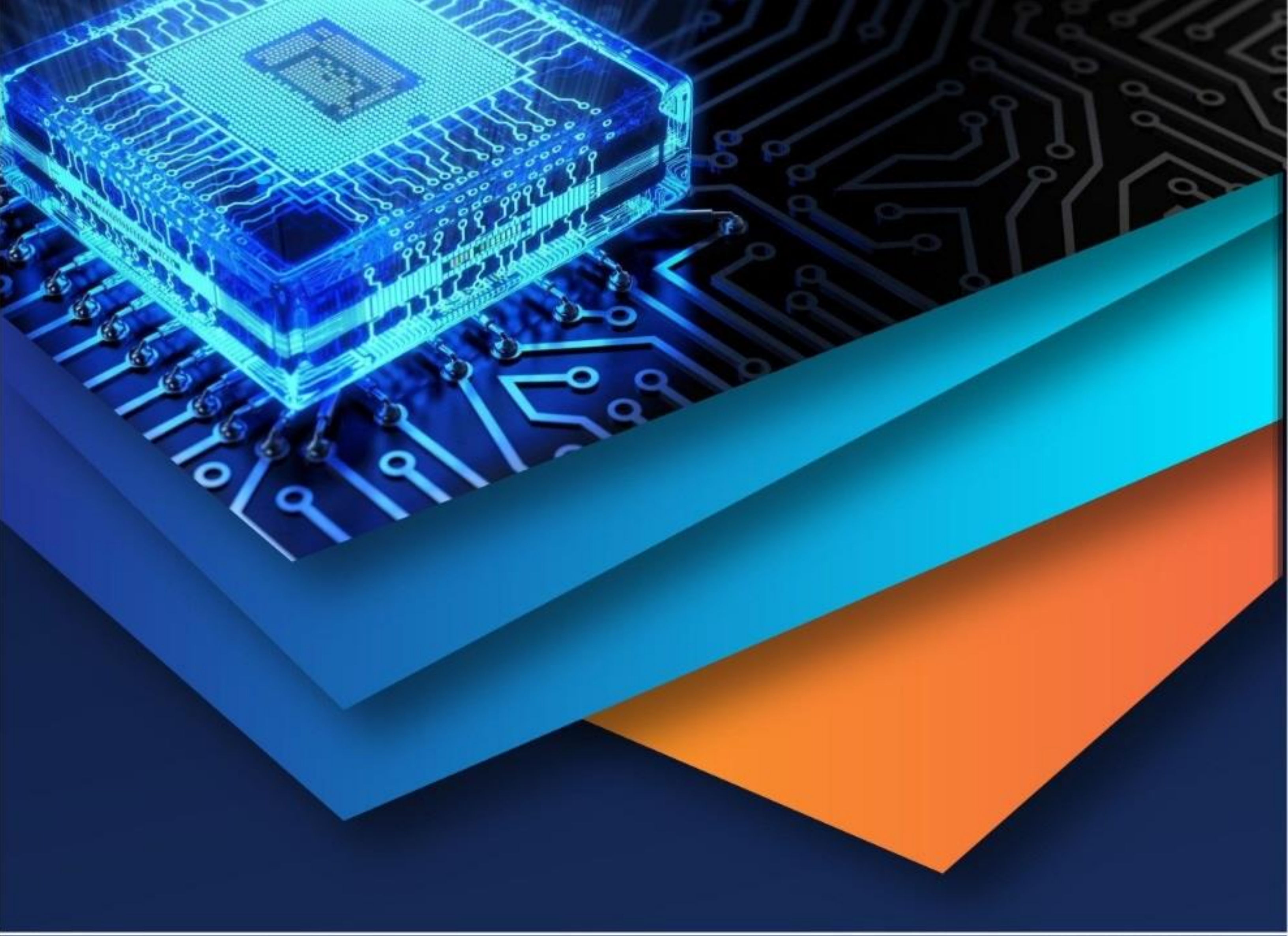
[6] Mr. Girme R. B, Prof. Marathe V. R. - Real Time Indian Sign Language Recognition System using SVM Classifier, International Journal of Advance Engineering and Research Development Volume 5, Issue 01, January -2018

[7] Mahesh Kumar N B - Conversion of Sign Language into Text, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161

[8] Joyeeta Singha, Karen Das - Recognition of Indian Sign Language in Live Video, International Journal of Computer Applications (0975 – 8887) Volume 70– No.19, May 2013

[9] Er. Vandana Soni, Dr. Tripti Arjariya - Survey Paper of Indian Sign Language Gesture System Weighted Euclidean Distance Based, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[10] Ms Kamal Preet Kour, Dr. (Mrs) Lini Mathew - International Journal of Technical Research & Science, Literature Survey on Hand Gesture Techniques For Sign Language Recognition, ISSN No.: 2454- 2024 (online)



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY