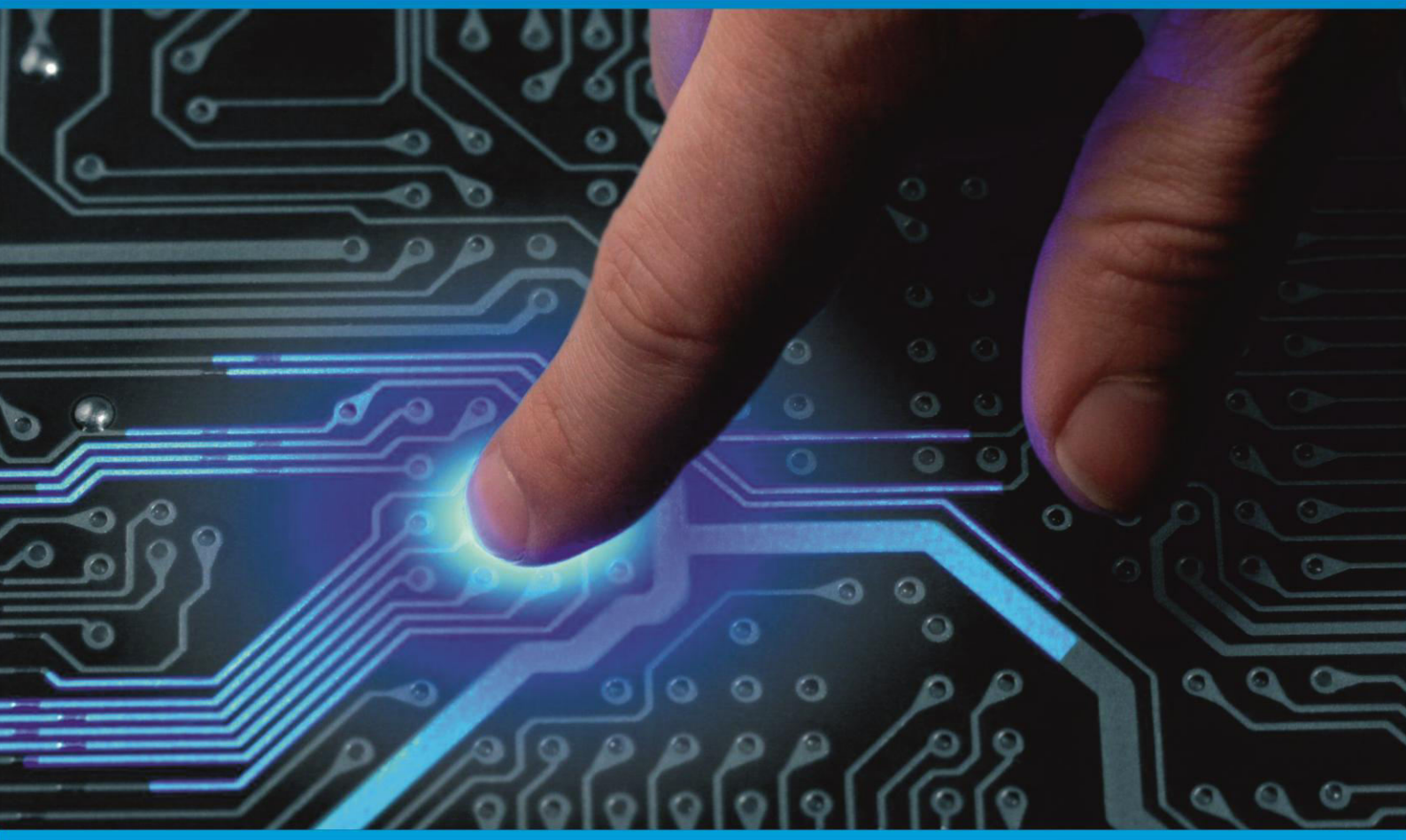
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

Interactive Application for Vocally Disabled People

Huda Noor Dean[1], Rafeekh AP[2], Prameela S[3]

Assistant Professor, Dept. of CSE., College of Engineering Trikarpiur, Kasarghod, CAPE, Kerala, India[1,2]

Assistant Professor, Dept. of CSE., College of Engineering Punnapra, Allapuzha, CAPE, Kerala, India[3]

ABSTRACT: Communication is treated as a life skill. All living beings like to communicate, to share and express their views and ideas to the community surrounding them. One of the important problems that our society faces is that people with vocal disabilities find it difficult to communicate. The access to communication technologies has become essential for the speech disabled people. They use Sign language for communication among themselves but common man may not sign language. In this paper we try to bring an interface system for the vocally disabled by using various technologies so as to ease the lives of people. We have used machine learning to implement the same by using python. The system creates a model which is trained by various gestures given by the users. Once the model generation is completed, the user could use the system by giving his/her gesture to the system and the system in turn will produce a voice signal which would be audible to the external and hence the communication is established.

KEYWORDS: sign recognition, machine learning, python, convolution neural network, recognition of gestures and speech output

I. INTRODUCTION

About 15% of the world's population lives with some form of disability and this 15% result to about 500 million which is not a small count. Out of this 500 million people, 6-8 million people are vocally impairment. The communication between a vocally impaired and hearing person poses to be an important disadvantage compared to communication between blind and ancient visual people. This creates an extremely little house for them with communication being associate degree elementary aspect of human life . The blind people can speak freely by implies that of ancient language whereas the dumb have their own manual-visual language referred to as language

Sign language is also very useful for the people who are either able to speak or hear which is a combination of different gesture, shape and movement of hand, body and facial expression. Each gesture or movement of hand, facial expression and body movement has special assigned meaning [1]. Aim of developing hand gesture recognition system is to establish an interaction between human and computer and hence recognize the hand gesture automatically which can be used later on for controlling robots or for recognizing the sign language[1] Hand gesture may be of static type or dynamic type. In static type certain position and posture is used for communication. In dynamic type of hand gesture, sequence of posture is used for communication. Real time application requires dynamic hand gesture recognition.

The interactive application for dumb people is also a tool that recognize the languages of dumb people and covert it in to the normal speaking language. This product also will help elderly people as they are another category of people who faces the same problem. The system is trained for personal signals and not a unified signed language. Each one will have their own set of signs for a word thus with proper training, the system could also help aged ones. Another set of people who would benefit out of this will be paralysed people. We expect to overcome the difficulties faced by these people in communication by lowering the barrier in communication with an application that uses image processing and machine learning. The device can act as an artificial tongue. The application can be used by the vocally disabled people for communication. Training and testing are the basic steps for this application.as the system is personalized one, training and testing should be performed separately for each individual. The basic principle behind this application is Machine learning. This system works in the manner that when the vocally disabled person shows a sign in front of this system, Then system converts the sign into corresponding words and sound. The primary goal of this project is to provide a standard lifestyle for dumb peoples as normal ones.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

II. RELATED WORK

In [2] Hand Gesture Recognition for Deaf People Interfacing J. Bermejo, and E. Casanova proposes a Digital vocalizer system that make use of sign language or gestures to make understand what he/she is trying to say but it is impossible to understand by hearing people. So,this system makes a simple prototype by taking some of those gestures and converting it into audio and visual form so that they can be understood by everyone. For that the system makes use of an Arduino UNO Board as Atmega 328 Controller board to interface all of the sensors and actuators.The basic concept of this prototype is Artificial Neural Network. Some sensors are placed on the hand of deaf people which convert the parameter like finger bend hand position angle into electrical signal and provide it to Atmega 328 controller and controller take action according to the sign.In [3] FakhreddineKarray, MiladAlemzadeh, JamilAbouSaleh, Mo NoursArab,Human-Computer Interaction proposes a system of two units. A subsystem which can be clipped on the collar of the shirt (as a lavalier microphone). This consists of input – condenser microphone and output – speaker sections and a camera to capture and record human faces for future reference with their voice modulation. And another main system is a portable device where the user sends and receives messages as text. Also the device acts as a mobile phone and the communication is performed as text chatting. The communications between these two units are via Bluetooth Nadia R. Albelwi and Yasser M. Alginahiproposed a real time sign language recognition system for arabic sign language [4]. Haar transform method is used to track hand in the video frames and the bounded hand region becomes the area of interest. Fourier transformation is used on the resultant images to transform it into the spectral domain to form the feature vectors and the classification is done using KNN method. In [5] hand shapes are described by a modification of the Histogram of Oriented Gradient (HOG) descriptor by Stephan Liwicki and Mark Everingham for recognizing the British sign language. A joint histogram over quantized gradient orientation and position and classification was done using Hidden Marcov Model. In [6] for Persian sign language recognition AzadehKianiSarkalehl et.al. uses discrete wavelet transform to extract features from the gesture images and a multilayered perceptron neural network for the classification of gestures. Al-Jarrah and Halawani [7] presented a method for automatic translation of gestures of the manual alphabet in the Arabic sign language. This system utilizes the feature values that comprise of some length measures which indicate the positions of the fingertips. Classification is done using a subtractive clustering algorithm and fuzzy inference system. A video based Indian sign language recognition system was developed by P. V. Kishore and Rajesh Kumar using wavelet transform and fuzzy logic [8]. They used wavelet based video segmentation technique to detect shapes of various hand signs. Shape features of hand gestures are extracted using elliptical Fourier descriptions and PCA is used to reduce the dimensionality of the feature space.

III. PROPOSED SYSTEM

A. *Design of Proposed System:*

The proposed system is an application which takes actions from the user and gives out voice messages as the required output. The system consists of three modules: real time hand tracking, training gesture and gesture recognition .The system is trained to detect each hand gestures made by the person which are then preprocessed and clustered into groups and each cluster deals with a specific message or word that is to be delivered. When the input is given, it is processed by an image processing system and the output of this system is passed to the machine learning model. With the trained examples the actual message is predicted. This is then converted into voice by text to voice mechanism thus giving out the expected outcome.

We used the technology of machine learning and an ample amount of image processing techniques in this project. As ML is the basic technique used, we did have to follow the steps involved in it. The steps involved were,

1. Defining Objectives

   Each and every project requires to be first researched upon. We have to learn the background of the project. This project is to be completed by identifying maximum words as possible with accuracy. we have used the best suiting algorithms for implementing this system. The backend of the project is Tensorflow with Keras. It is the Python deep learning API. By using this softwares, we develop a system which decreases the communication gap between the vocally disabeled people and normal people.It's also useful for the old people.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

2. Data Gathering

   The quality and quantity of your data dictate how accurate our model is. The outcome of this step is generally a representation of data which we will use for training. Using pre-collected data, by way of datasets from kaggle,UCI,etc.

3. Preparing Data

   Wrangle data and prepare it for training. Clean that which may require it(remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data. In data cleaning procedure, we cropped the user input to the required size and the region of interest was cultivated. The image was then converted to grayscale image and calculated for their threshold values.

4. Data Exploration

   Visualize data to help detect relevant relationships between variables or class imbalances, or perform other exploratory analysis.Split into training and evaluation sets.Different algorithms are for different tasks; choose the right one.

5. Building a Model

   The goal of training is to answer a question or make a prediction correctly as often as possible. The concept of supervised learning is used in building this system. We used the algorithm Convolutional Neural Network along with GestureNET so as to get a model that could classify the images with utmost accuracy. The structure is designed by the json file which defines the scaling size and mapping indices.An epoch of 75 was used to train the model. We have also used the scikit-learn,keras and the pyimagesearch packages to implement the system.

6. Model Evaluation

   Uses some metric model or combination of matrices, to “measure” objective performance of the model. Test the model against previously unseen data. This unseen data is meant to be somewhat representative of model performance in the real world, but still helps tune the model.

7. Prediction

   Using further data which have, until this point, been withheld from the model, are used to test the model; a better approximation of how the model will perform in the real world.

B. *Block Diagram of Proposed System:*

The block diagram shows that, First we create the data set which is used as the training information. The training information is stored in a database. The database is clustered on the basis of different words. Each set will have N number of collected pictures. These will then be used for training the system. The training set is grouped into clusters based on the different features. After the training is done the model is created.

If the user gives an input, the model processes the input by using some part of image processing and compares the input into the dataset that we already saved. If the model identifies the input it converts the input into the text format and then its voice message as the output to the user.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

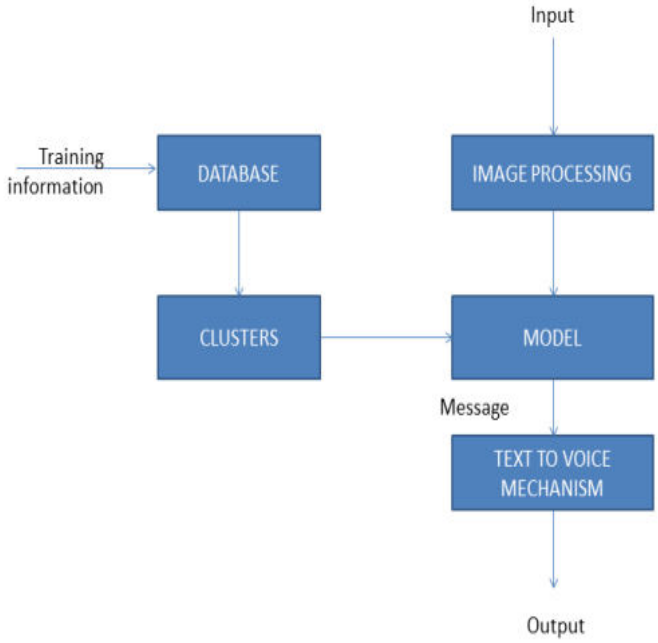


Fig.1. Block Diagram of Proposed system

1. Data Set creation module

Since we are building this project from the bottom, the first thing we need to do is to create the data that we are going to use for training the Neural Network model. The project uses a separate module named "roi.py" for the data creation i,e for capturing the pictures. In this module we use the image processing techniques to extract the details from the images captured. Each image is resized, and converted into a grayscale image. This is then passed through activation functions so as to get the threshold values.

Fig.2. Dataset_good

Fig.3. Dataset_hai

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

Fig.4.Dataset_morning

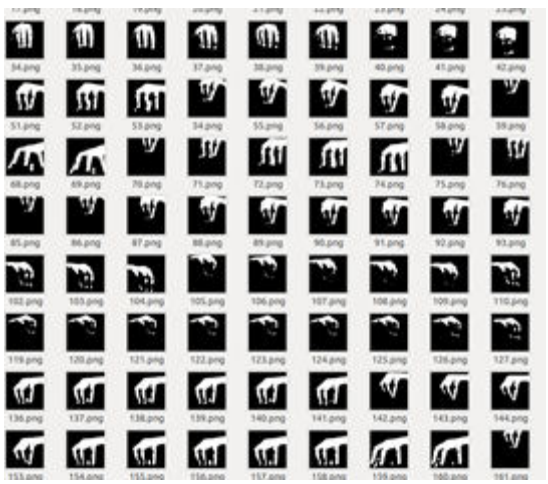


Fig.5 .Dataset_night

2. Training module

The system is trained using the supervised learning algorithms. Supervised learning is a technique in which we teach or train the machine using labelled data. For training first we grab the list of images in our dataset directory , then initialize the images. Once we load the image by using a camera ,it will create a lot of noise. It has different pixel colors. So we take the part of the picture and also extract the details which we want. So that first we convert the image into grayscale so we can remove the nose. For example if we give “hai” as an input it will only concentrate the region of the palm and crop that area and turn into a black and white portion. In our dataset 75% of the data are used for training and the remaining 25% of the data are used for testing purposes. For the designing phase, we used Convolution Neural Network algorithm as the base algorithm.It is a specialized type of neural network model designed for working with 2D image data It is the main categories to do image recognition, image classification, object detection, recognition faces etc. It takes an input image, then processes the image and classifies it under certain categories like size, labels, colors etc. Computer sees the input image as an array of pixels and it depends on image resolution.

To train and test, the CNN model passes each input image through a series of convolution layers and apply some function to classify an object with probabilistic values between 0 and 1.

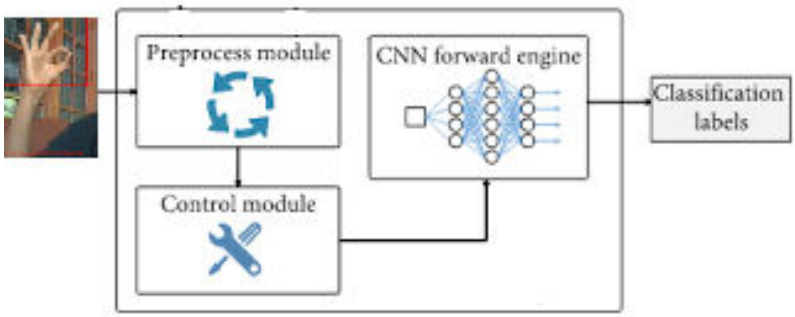


Fig.6. Training using CNN

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

3. Text to voice message module

This module uses various activation functions for the proper prediction of the output. We used algorithms such as Conv, ReLU, Maxpool, Dense, Normalization, SoftMax as activation functions. We use e - speak for text to voice conversion, e - speak is a synthesizer used to convert text to voice. ReLU stands for rectified linear unit and is a type of activation function. It is defined as

$$Y = \text{Max}(0,x)$$

Max pooling is a sample - based discretization process. The objective is to down - sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. The softmax activation is normally applied to the very last layer in a neural net, instead of using ReLU, sigmoid, tanh, or another activation function. The reason why softmax is useful is because it converts the output of the last layer in your neural network into what is essentially a probability distribution. A Dense layer feeds all outputs from the previous layer to all its neurons, each neuron providing one output to the next layer. It's the most basic layer in neural networks. A Dense has ten neurons

IV. EXPERIMENTAL RESULTS

The proposed system as an application has been trained with the given set of datasets. It took input from user as real time, and it was preprocessed and classified using CNN to get the desired result with accuracy. Voice messages as per output were also produced.

Fig.7. "good"

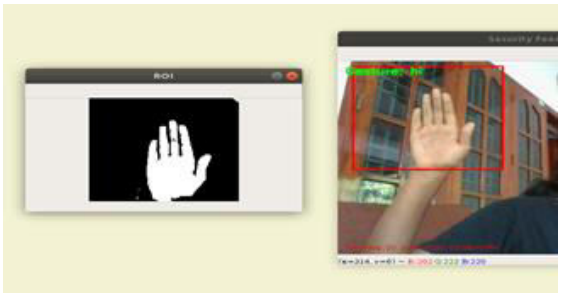


Fig.8. "hai"

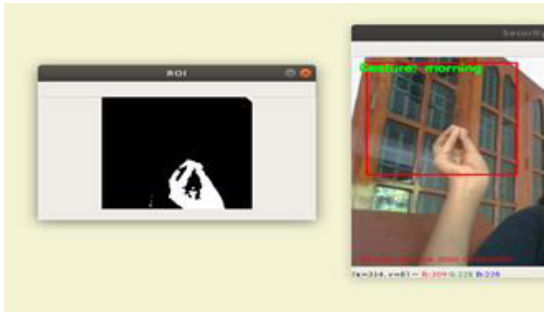


Fig.9. "morning"

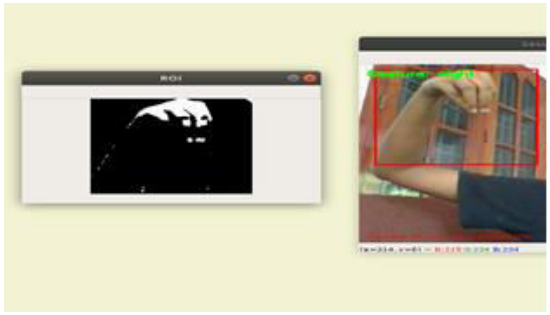


Fig.10. "night"

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

In this graph X axis indicates the number of epochs and Y axis indicate the loss or accuracy. The red and green line indicate the training and validation loss and blue and black line indicate the training and validation accuracy.We are trying with a data set of 150 images with an epoch of 75. We have actually got an accuracy level of 70%.

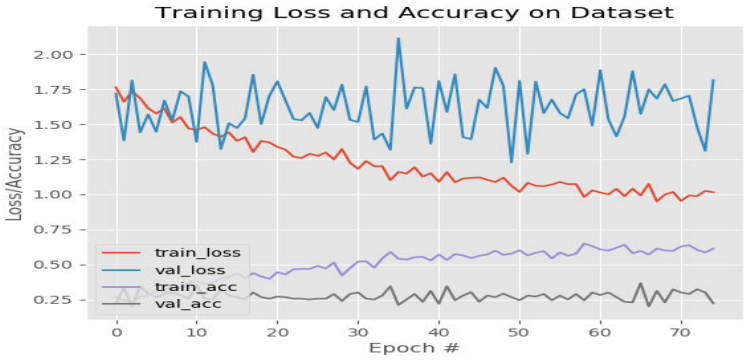


Fig.11. Training loss and accuracy on dataset

V. CONCLUSION AND FUTURE WORK

By using this application dumb person can easily interact with normal person, This system gives sound as output so it will be easily understand by others. The system aims to lower the communication gap between dump people and normal world, since it facilitates two way communications. The system overcomes the necessary time difficulties of dumb people and improves their manner.

Future scope

- The large data input to be enabled.
- Embed the system so as to get an application.

Limitations

- The system requires a large amount of memory capacity.
- It trained only for a few data sets.
- As there is a huge amount of memory, the chance of making a mobile application is difficult, so there is risk of carrying the system wherever we go.
- This system is a personalized one. So each individual must be trained and tested separately.

REFERENCES

[1] Hand Gesture Recognition for Sign Language Recognition: A Review IJSTE -International Journal of Science Technology & Engineering | Volume 2 | Issue 10| April2016 SwapnilChandelMrs. AkankshaAwasthi
[2] Incertis, J. Bermejo, and E. Casanova, “Hand Gesture Recognition for Deaf People Interfacing,” The 18th International Conference on Pattern Recognition, 2006 IEEE.
[3] FakhreddineKarray, MiladAlemzadeh, JamilAbouSaleh, Mo Nours Arab, (2008) .“Human-Computer Interaction:Overview on State of the Art”, International Journal on Smart Sensing andIntelligent Systems, Vol. 1(1).
[4]Nadia R. Albelwi, Yasser M. Alginahi, “Real-Time Arabic Sign Language (ArSL) Recognition“ International Conference on Communications and Information Technology 2012
[5]Stephan Liwicki, Mark Everingham, "Automatic Recognition of Fingerspelled Words in British Sign Language", School of Computing University of Leeds
[6]AzadehKianiSarkalehl, FereshtehPoorahangaryan, BahmanZan, Ali Karami, “A Neural Network Based System for Persian Sign Language Recognition” IEEE International Conference on Signal and Image Processing Applications 2009.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904189|

[7]Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence 133 (2001) 117–138.

[8]P. V. V. Kishore and P. Rajesh Kumar, "A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic", IACSIT International Journal of Engineering and Technology, Vol. 4, No. 5, October 2012

[9]R. Rokade, D. Doye, and M. Kokare, "Hand Gesture Recognition by Thinning Method" International Conference on Digital Image Processing (2009).

[10]R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi. "Exploiting depth discontinuities for vision-based fingerspelling recognition".In IEEE Workshop on Real-time Vision for Human-Computer Interaction, 2004.

[11]M. Lamari, M. Bhuiyan, and A. Iwata. "Hand alphabet recognition using morphological pca and neural networks". In Proc. IJCNN, pages 28392844, 1999.

[12]A. Park, S. Yun, J. Kim, S. Min, and K. Jung. "Real time vision-based Korean finger spelling recognition system". Proc. World Academy of SET, 34:293298, 2008

[13]P. V. V. Kishore and P. Rajesh Kumar, "A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details