iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5277

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

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 V May 2020- Available at www.ijraset.com

The Silent Voice - Communication Interface for Deaf and Dumb People

Chahat Sharma[1], Aditi Bajpai[2], Akarshita Chopra[3], Shubhangi Singh[4], Utkarsh Tomar[5]
[1, 2, 3, 4, 5]*Inderprastha Engineering College, Ghaziabad(U.P.), India*

Abstract: ***Hand gesture is a channel of communication between dumb and deaf people. Hand gestures are physical movement by using eyes and hands and non-physical movement is facial appearance, head movement, body position etc. Contemporarily, there are innumerous techniques for capturing and recognizing the hand gestures.***
Substantial literature exists and extensive analysis, groundwork, exploration and inquisition has been done till now regarding gesture identification systems to resolve communication problem with the disabled (visual and hearing impaired) people and how it can be used with respect to real time.
This paper presents the method of developing such a system using latest learning techniques such as CNN and its scope of implementing it for because not every single typical person can comprehend the gesture based communication. Paper also discusses about certain areas and the future research which can be done in regard to CNN for efficient application of the technique in studying the complexities prevailing in the usage of the system.
Keywords: Convolutional Neural Networks(CNN), American Sign Language(ASL), Artificial Neural Network(ANN), Histogram of oriented gradients(HOG), Static Hand Gesture Recognition

I. INTRODUCTION

Communication is imparting, sharing and conveying facts, bulletins, recommendations and feelings. Of them, sign language is the mpst frequently used ways of non-verbal communication which is gaining impetus and strong foothold due to its applications in a various number of fields. The most prominent application of this method is its usage by differently disabled persons like deaf and mute people.

They can be in contact with the non-signing persons without the help of a translator or interpreter by this method. Some other applications are in the automotive sector, transit sector, gaming sector and also while unlocking a smartphone [1]. The sign identification can be done in dyad ways: consistent(static) gesture and dynamic gesture [2]. While communicating, the static gesture makes use of hand shapes while the dynamic gesture makes use of the movements of the hand [2]. Our paper focuses on static gestures.

But the human hands have very complex articulations with the human body and therefore a lot of errors can arise [3]. Thus it is tough to recognize the hand gestures. Our paper focuses on detecting and recognizing the hand gestures using different methods and finding out the accuracy by those methods.

Also we see the performance, convenience and issues related with each method. Currently a lot of methods and technologies are being used for sign and gesture recognition.

Among them the most common ones used are Hand Glove Based Analysis, Microsoft Kinect Based Analysis, and Convolutional Neural Networks(CNN).

One of the objectives of these methods is to keep the head above water between speech and hearing impaired people with the normal people and also successful and smooth integration of these impaired(visual and hearing) people in our society. In our research paper we build a real time communication system using the advancements in Machine Learning.

Currently the systems in existence either work on a small dataset and achieve stable accuracy or work on a large dataset with unstable accuracy.

We try to resolve this problem by applying Convolutional Neural Network (CNN) on a fairly large dataset to achieve a good and stable accuracy. Uses. Hand gesture recognition is a way of understanding and then classifying the movements by the hands.

II. LITERATURE SURVEY

In order to keep one going between hearing and speech impaired members, different approaches have been used by researchers for recognition of various hand gestures. These techniques can be majorly split into three categories - Hand Segmentation Approach, Gesture Recognition Approach and Feature Extraction Approach. Two categories of visual-based hand gesture recognition can be used. The first one is a 3-D hand gesture model that works by comparing input frames which makes use of sensors like gloves, helmet, etc. [9]. The other one is Microsoft Kinect based analysis which makes use of Kinect cameras. Kinect hardware gives accurate tracking of several user joints. So a huge dataset is required for the 3-D hand gesture model since it requires a huge data set and also has a higher hardware cost due to sensors on the gloves. This glove based model for American Sign Language was proposed by Starner and Pentland [10].

It is not practically possible for the user to everytime wear gloves. The 2-D hand gesture model makes use of an image dataset for feature extraction and detection. There are many other approaches used for image based gesture recognition like ANN (Artificial Neural Network), HMM (Hidden Markov Model), Eigenvalue based and Perceptual color based. The characteristic/attribute vectors drawn out from the image are inputted into HMM [11]. For classification, particle filtering and segmentation methods like Support Vector Machine (SVM) are used where the image frame is converted into HSV color space as it is less sensitive to light effects [12]. Feature extraction can be employed using various methods. One of the most used methods for feature extraction is by Contour Shape Technique which extracts the boundary information of the sign.

The Table I on the next page summarizes the various techniques. It presents various merits and issues found in respective literature in relation to the techniques used for the research work.

III. CURRENTLY USED METHODOLOGIES

1) *Feature Extraction*- A feature is a distinctive attribute of one or more quantifications computed so that it measures some distinct significant characteristic of the object or entity [5]. It is a special and unique form to reduce the number of random variables to consider i.e.,dimensionality reduction. In pattern recognition and also in image processing, if the input given is quite large for processing, then it is doubted to be redundant and eventually the input data which is given will transform into a reduced delineation set of features or distinctive attributes [4]. Feature extraction can be defined as a process of transforming input data in a set of features. The general expectation is that these distinctive attributes will extract the piece of information which is relevant from the dossier if we extract the features carefully for the purpose to enact the required task using this reduced portrayal instead of the entire dossier. Some issues with feature extraction are, firstly the distinctive features should carry ample knowledge about the image and should not require any cognitive content for their extraction [5]. Secondly, the features must be facile to enumerate for the purpose to make feature extraction more realistic for a cosmic image collection and swift recoup.
2) *Hand Segmentation Approach*- Hand tracking and Segmentation should be always done in an efficient manner as they are the keys of success towards any gesture recognition, because of the challenges vision based methods pose such as intensity of the continuous variation in lightning, many objects in the background(complex) and detection of the skin color. Color is a very powerful descriptor for object detection. Thus, hue knowledge was used for the segmentation purpose, which is unwavering to rotation and geometric variation of the hand [6]. Humans see color component's features such as saturation, hue and the brightness component more than the rate of basic colors which are RGB(red, green and blue) [6]. These color models represent the standardized way of a particular color. It is a space-coordinated system in which any color which is specified is regarded as a single point. Here, using discrete hue spaces for robust hand detection and segmentation, three techniques were introduced. Hand tracking and segmentation (HTS) technique using HSV color space is identified for the pre-processing of the HGR system. Some issues with hand segmentation are, firstly some objects, which are irrelevant, might overlap with the hand. Also, performance of the hand segmentation algorithm is degraded when the distance between the end user and the web camera is more than 1.5 meters [7]. Lastly, hand segmentation restricts the user to make some gestures in a particular manner, like gestures must be made with the right hand only, the arm should be vertical, the palm should face the web camera and the backdrop should be clear and uniform.
3) *Glove based hand gesture recognition*- Glove based approaches make use of gesture or capacitive touch sensors embedded into gloves to recognize hand gesture. The widely used methods make use of hand motion to convey hand signs and the motion is tracked and translated to text. Hand motions are categorized using clustering techniques such as k means. Other approaches use charge-transfer touch sensors for translation by using On / Off binary signals. These approaches achieve high accuracy but incur high cost due to the necessary hardware.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

IV. PROPOSED METHODOLOGY

This section provides the description of the dataset and CNN configuration that was used. The flowchart of methodology is shown on Figure 1. The procedure is the amalgamation of statistics collection, pre-processing, configuring the CNN and building the prototype.

1) *Input Details and Training Data-* Images needed to instruct and affirm the prototype were collected using a web camera. The signs were executed by various people before the web camera. It is taken for granted that the input images included exactly one hand, gestures/signs were made with the right hand of the user, the palm facing the web camera and the hand was roughly vertical. The recollection/recognition process will be less complex and more efficient if the background is less complex and the contrast is high on the hand. So, it is taken for granted that the background of the images was less complex and uniform.

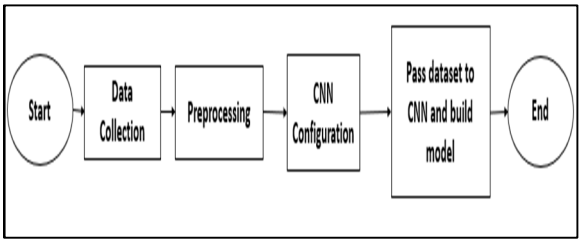


Figure 1 System Framework

2) *Histogram of oriented gradients-* HOG is a lemma used in visual systems for object/entity detection or classification [8]. This approach counts the occurrences of gradient orientation in localized portions of an image. HOG is calculated on a concentrated mesh of evenly spaced cells and it uses intersecting local contrast normalization for more refined and better accuracy. In an effort of extracting the HOG features, the hue image relating to the tracked hand is in the first place resized to fit 64x128 pixels, then divided into 16x16 blocks with a 50% intersection. This leads to 105 (7x16) blocks in total. Each representing 2x2 cells with a size of 8x8 as shown in fig. finally the resulting gradient orientations vector is quantified using 9 bins.

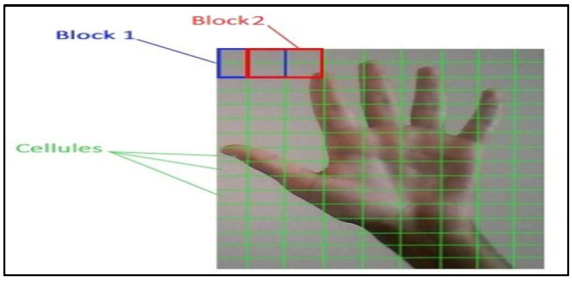


Figure 2 Example of dividing an image into blocks and cells using HOG algorithm.

3) *Dataset-* We selected around 44 static gestures (One, Two, Three, Palm Opened, Palm, Closed, OK, Zero, All the best etc.) for identification. Each class includes 2400 images for training purposes. So the total count of images is around 1,05,600 for training.

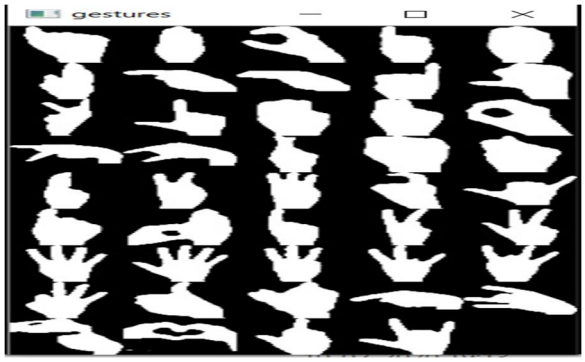


Figure 3 Specimen Images from self-guided Dataset

4) *Image Recognition-* In order to achieve higher accuracy as compared to existing systems and to keep the system computationally lightweight, we use Convolutional neural network (CNN) for image recognition. Convolutional neural networks are a class of feed-forward artificial neural networks commonly used for visual analysis tasks. They comprise neurons which act as learnable parameters having their own weight and biases. The entire neural network learns with the help of a loss function a step-size is used to fine tune the learning. The input layer executes the information which gets propagated through the various layers and an output is generated, the generated output is compared with the actual output and the system updates its weights and biases to correct itself, this step is crucial and is known as backpropagation and this process done iteratively is called as training. The training duration of CNN is decided according to the size, the number of layers and also the learning rate. The CNN was trained using the ASL sign language image dataset consisting of around 35K images with each class having a minimum of 1200 images. The dataset consisted of gray scale static sign images concerning alphabets and numbers. The character labels associated with each image were converted into binary vectors using one hot encoding, thus converting categorical values into numbers. The proposed architecture of our CNN includes three convolutional layers with 32, 64, 128 number of filters, having intermediate max-pooling layers and Rely activations.

A kernel of size 3 and pool size of 2 was used accordingly. The last three layers consisted of a flattening layer and fully connected layers with dropout layers in between in order to avoid overfitting. The final dense layer is of size 36 corresponding to the number of class labels with softmax activation. The code to the CNN will be the gray scale processed image resized to 28 * 28 as per the dataset and the yield of the CNN would be a probability distribution to classify the image into probabilistic values ranging between 0 and 1. The loss function used for training was categorical cross entropy and the optimizer used was rmsprop. The training was conducted for 250 epochs with a collection of 512. For any image fed into the CNN, it outputs a probability distribution. The node containing the highest probability value is considered as the output node and the correct label against that node is outputted. In this way the system determines what sign the user performed.

V. RESULT

A hybrid segmentation which included the feature fusion-based sign word recognition system was brought to light in this paper. For this aim, a model was proposed that was able to translate sign gestures into text. The proposed model included image preprocessing, feature extraction and fusion, and gesture classification. The outcomes clearly depicted that in the present environment, approximately 97.28% precision is attained using trained features of CNN.

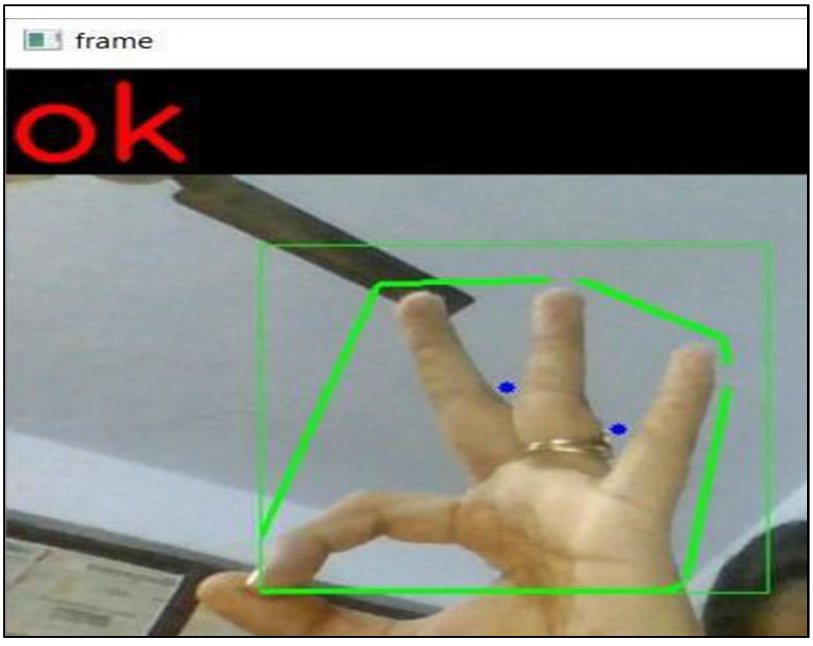


Figure 4 Gesture Recognized by The Software

VI. FUTURE SCOPE

A. Noise reduction for more accurate result.
B. Use of Graphics Progressing Unit (GPU).

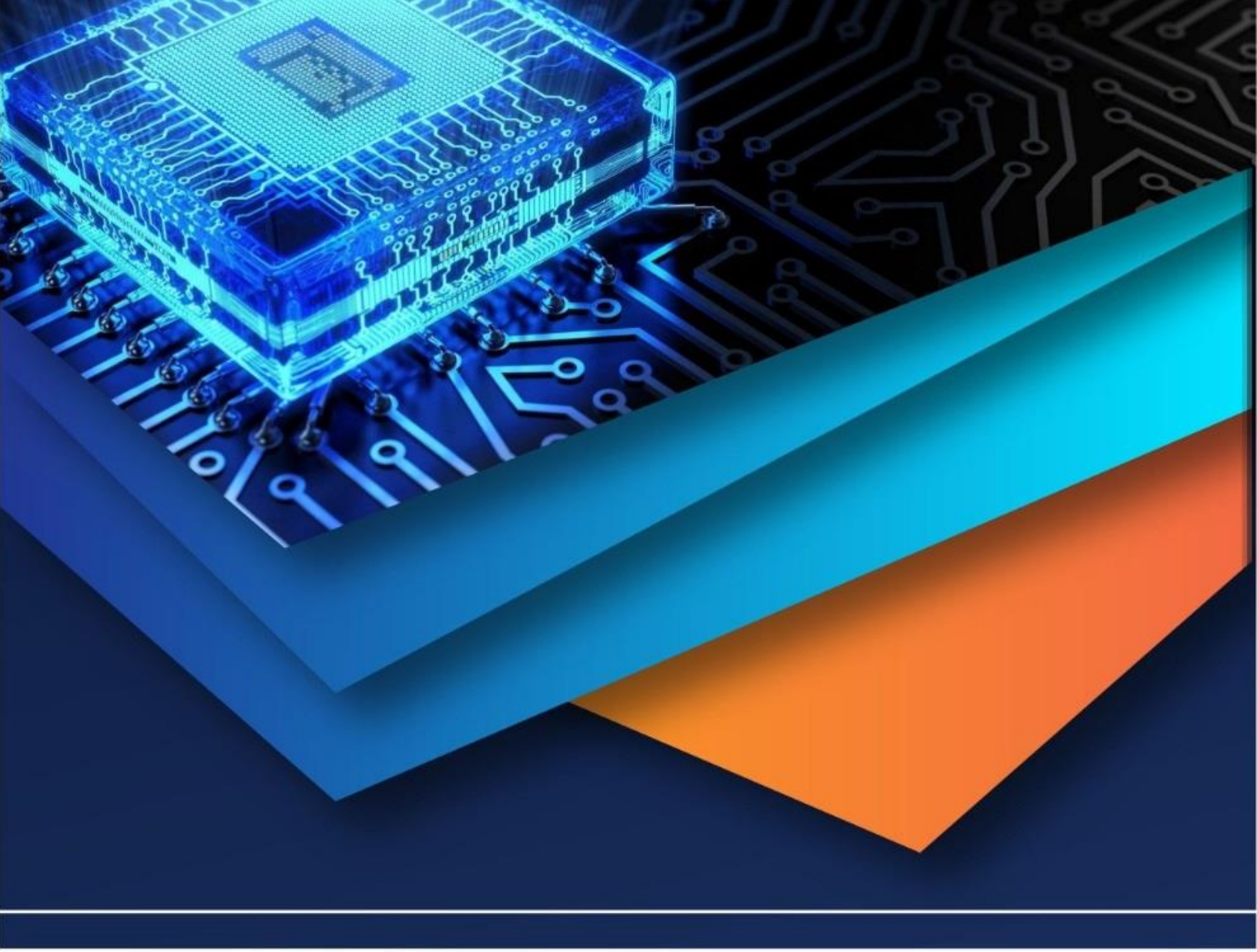
REFERENCES

[1] Gesture Recognition (2018, October, 4) Wikipedia [Online] Available:
[2] Priyanka C Pankajakshan, Thilagavathi B, Sign Language Recognition System, IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS15.
[3] Jobin Francis and Anoop B K. Article: Significance of Hand Gesture Recognition Systems in Vehicular Automation-A Survey. International Journal of Computer Applications 99(7):50-55, August 2014.
[4] Sanaa Khudayer Jadwaa, Feature Extraction for Hand Gesture Recognition: A Review, International Journal of Scientific Engineering Research, Volume 6, Issue 7, July-2015
[5] George Karidakis et al Feature Extraction-Shodhganga
[6] Archana Ghotkar, Gajanan K.Kharate, Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction, International Journal of Human Computer Interaction
[7] Rafiqul Zaman Khan, Noor Adnan Ibraheem, Comparative Study of Hand Gesture Recognition System, SIPM, FCST, ITCA, WSE, ACSIT, CS IT 06, pp. 203213, 2012.
[8] N. Dalal, N. and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, San Diego, CA, USA.
[9] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden Markov models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.
[10] Camastra, Francesco, and Domenico De Felice. "L VQ-based hand gesture recognition using a data glove." Neural Nets and Surroundings. Springer BerlinHeidelberg, 2013. 159-168.
[11] Lang, S., B. Marco and R. Raul. Sign Language Recognition Using Kinect. In: L.K. Rutkowski, Marcin and R. T. Scherer, Ryszard Zadeh, Lotji Zurada, Jacek (Eds.), Springer Berlin / Heidelberg,pp:394-402,20 11.
[12] V. K. Verma, S. Srivastava, and N. Kumar, "A comprehensive review on automation of Indian sign language," IEEE Int. Conf, Adv. Comput. Eng. Appl. Mar 2015 pp. 138-142

Table I. Comparison of literature review

S.N.	Author(s)	Year	Methodology	Dataset	Tools	Merits	Issues
1.	Priyanka C Pankajakshan, Thilagavathi B	2012	Image capturing using webcam and classification using Artificial Neural Network	Database of 25 images for 5 types of gestures captured using webcam.	Desktop PC, webcam, speaker	Can be implemented in a hardware which supports image processing applications.	Use of only static gestures, no use of dynamic gestures.
2.	Sanaa Khudayer Jadwaa	2015	Feature Extraction and classification	Not Specified	Desktop PC, Webcam	Use of different hand features for the recognition process that make the recognition process more accurate.	More precise research to realize the ultimate goal of humans interfacing with machines
3.	Rafiqul Zaman Khan, Noor Adnan Ibraheem	2012	Segmentation, Feature Extraction, Classification	American Sign Language	Desktop PC, Webcam	Use of HMM for dynamic gesture is perfect and its efficiency especially for robot control.	Performance of the algorithm decreases when distance is greater than 1.5 meters between user and camera.

4.	N. Dalal, N. and B. Triggs	2005	Feature Extraction and SVM Classification	509 training and 200 test images of pedestrians in city scenes	Desktop PC, Webcam	HOG in a dense overlapping grid gives very good results for person detection	Use of HOG descriptors lowered the speed.
5.	T. Starner and A. Pentland	1995	Feature Extraction, Training using HMM network	American Sign Language	Desktop Pc, video camera	Through HMM low error rates were achieved on both the training set and an independent test set.	Small training set
6.	Camastra, Francesco, and Domenico De Felice	2012	Feature Extraction, Classification using LVQ	7800 right hand gestures	Desktop PC, Camera	Average time of 140 CPU ms to recognize a gesture.	Use of external hand glove
7.	Lang, S., B. Marco and R. Raul	2012	Feature Extraction and Classification using CNN	American Sign Language	Desktop PC, Kinect Camera	Kinect depth cameras allow offer 3D data without a complicated camera setup and efficiently extract the users' body parts,	Use of GPU



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY