

NAAN MUDHALVAN – IBM CLOUD

BASED PROJECT INNOVATION

DOCUMENT

PHASE 2 :

TEAM:7

1)K.RAJESH(T.L) - 422421106023

2)A.ALWIN XAVIER - 422421106003

3)T.DHIVAGAR - 422421106009

4)E.SURYA PRAKASH - 422421106029

Sign Language and Gesture Recognition for Image Recognition

Abstract - Every day we see many people who are facing illness like deaf, dumb and blind etc. They face difficulty to interact with others. Previously developed techniques are all sensors based and they didn't give the general solution. This paper explains a new technique of virtual talking without sensors. An image processing technique called Histogram of gradient (HOG) along with artificial neural network (ANN) has been used to train the System. Web Camera is used to take the image of different gestures and that will be used as input to the Mat lab. The software will recognizes the image and identifies the cores pending voice output which is played using voice replay kit. This paper explains two way communications between the deaf, dumb and normal people which means the proposed system is capable of converting the sign language to text and voice.

Key Words: Histogram of gradient (HOG), Artificial Neural Network (ANN), Region Of Interest (ROI), Digital Image Processing, Flex Sensors, Orientation Binning.

1. INTRODUCTION

The challenge faced by dumb and deaf people while communicating with the system in work place , since they cannot hear it, dangerous to go places alone because they cannot hear car, bikes, or other people coming. They cannot adapt to the surrounding environment quickly and respond to other normal people and expressing themselves is hard. The record history of sign language in western societies starts in the 17th century as a visual language or method of communication. Sign language is composed of a system of conventional gesture, mimic, hand sign and figure spelling, plus the use of hand position to represent letters of the alphabet. Sign can also represent complete idea or phrase. The main purpose is to provide speech and text output using hand gesture sign language without using any sensor for dumb people in smart way.

1.1 Problem statement

Dumb people use hand signs to communicate, hence normal people face problem in recognizing their language by signs made. Hence there is a need of the systems which recognizes the different signs and conveys the information to the normal people.

1.2 Aim

Aim of our project is to develop a concept of virtual talking system without sensor for people who in need, this concept achieving a by using image processing and human hand gesture input. This mainly helps to people who can't talk with other people.

2. EXISTING SYSTEM

In existing system the module was developed for dumb person using flex sensor, there user hand is attached with the flex sensors. On this module the flex sensor reacts on bend of each finger individually. By taking that value controller starts to react with speech, each flex sensor holds unique voice stored in APR Kit and for each sign it will play unique voice. And in other existing system, the work is done only for some alphabets and not for the words or sentences, and accuracy obtained is very low.



Fig- 1: Usage of flex sensors gloves

2.1 Limitations of existing system

- In existing system it's restricted to only 10 voice announcements it may reduce product capacity
- One of the major problem of the existing system is Dumb person should always carry the hardware with him
- User can't do any other work with flex sensor on fingers and also sensors should be placed straight
- The controller may think that the user is giving command and finally it may result in unwanted results and less hardware lifetime

3. PROPOSED SYSTEM

In the proposed system the unable or dumb person should provide a gesture or sign image to the system. The system evaluates the sign input with IBM cloud image processing technique and classifies the input to the recognized identification. Later it initiates the voice media through the system when the input image matches with the given dataset. And the output will be shown in the text format too. This is a prototype to develop the concept of converting the sign language to speech and text. The aim of this paper is to provide an application to the society to establish the ease of communication between the deaf and mute people by making use of image processing algorithm.

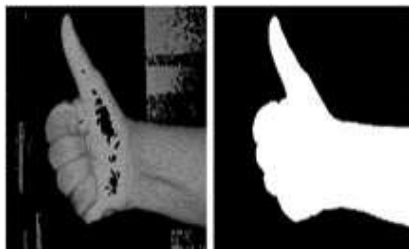


Fig-2: Using IBM cloud to recognize sign language

3.1 Advantages of proposed system

- When comparing with existing system user can give more signs
- The module provides two way communications which helps in easy interaction between the normal people and disables
- Easy to Interface
- Flexible

4. METHODOLOGY

4.1 Histogram of Oriented Gradients (HOG)

The histogram of oriented gradients (HOG) is a feature descriptor used in **computer vision** and **image processing** for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an **image**. This method is similar to that of edge orientation histogram but differ in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization or improved accuracy. The use of HOG is that local object appearance are sharp within an image can be described by distribution of intensity gradient or edge direction. The image is divided into small connected region called cells and for the pixels within each cells a histogram of gradient direction is compiled. [7] The descriptor is the concatenation of these histogram. Fig-3 shows the basic flow

The key advantages of HOG descriptor are:

Since it operate on local cells it is invariant to geometric and photometric transformation except for object orientation.

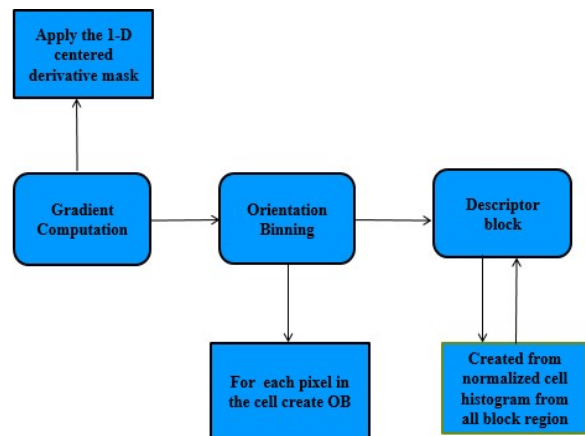


Fig-3: Basic flow of HOG

4.2 ARTIFICIAL NEURAL NETWORK (ANN)

ANN in image processing is used for gesture recognition in our method we use neural network to recognize different hand gesture, neural network is a connection of interconnected nodes with 3 layers input hidden and output layer. The given input propagates through the network to provide the required output. First we have to train the neural network to recognize different hand gesture so we give 100's of sample images as different input to obtain the different output. [7] The image of hand gesture are complex even though there are in grey scale so image are converted into binary image hence the resolution also reduces to 20*30 so the no of input given to network is only 600. Fig-4 show the basic flow of ANN.

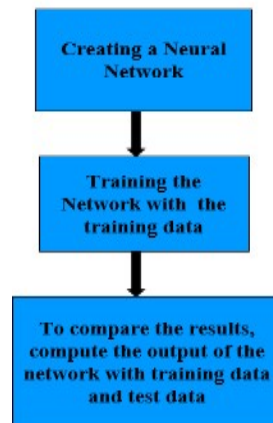


Fig-4: Basic flow of ANN

4.3 Region of interest(ROI)

Region of Interest, are samples within a dataset identified for a particular purpose. The steps involved are:

- ROI filtering- defines a mask to separate the ROI from the background and then applying filter to ROI only
- Create a binary mask-classifies image pixels either it belongs to ROI or background
- Sharpen region of interest in an image-increasing the sharpness of a specific ROI
- Apply custom filter to region of interest in image-defining functions and masks to filter a ROI

4.4 Digital Image Processing

Image refers to a two-dimensional intensity function $f(x, y)$, where x and y denotes spatial coordinates and the value of f at any point (x, y) is proportional to the brightness of the image at that point. [2]

A Digital image is a representation of a two dimensional image as a finite set of digital values, called picture elements or pixels. [2]

Digital image processing focuses on two major tasks

- Improvement of pictorial information for human interpretation
- Processing of image data for storage, transmission and representation for autonomous machine perception

There are mainly three levels of processing the image as shown in below figure

Low Level Process	Mid Level Process	High Level Process
Input: Image Output: Image	Input: Image Output: Attributes	Input: Attributes Output: Understanding
Examples: Noise removal, image sharpening	Examples: Object recognition, segmentation	Examples: Scene understanding, autonomous navigation

Fig-5: Digital image processing

4.5 Histogram

Histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. [2]

For an 8-bit grayscale image there are 256 different possible intensities, so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those grayscale value.

5. SYSTEM OVERVIEW

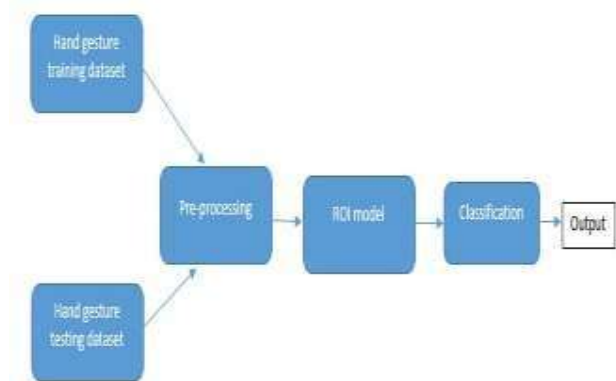


Fig-6: System overview

Here the hand gesture training dataset will take the input images and pre-processing is done. After pre-processing phase is done, the result is stored in the region of interest (ROI) model. In the ROI model the region that is required for classification is removed by applying the algorithms. The model is trained for different hand gestures and stores the data in the classification model. Later the testing is done for input images given by the user. Then the pre-processing is done for testing image using ROI, if the encoded input matches with the testing image then classification is done with the training images. If the input matches with the training data then respected output voice is played.

6. CONCLUSION

The sensor less sign language and gesture recognition system is a module which provides an easy and satisfactory user communication for deaf and dumb people. The module provides two way communications which helps in easy interaction between the normal people and disables. The system is novel approach to ease the difficulty in communicating with those having speech and vocal disabilities. The aim is to provide an application to the society to establish the ease of communication between the deaf and mute people by making use of image processing algorithm. Since it follows an image based approach it can be launched as an application in any minimal system and hence has near zero-cost.