

A.I Based Hand Gesture Speaking System For Dumb People

Mayur Dighe¹, Manthan Jadhav², Abhishek Kote³, Pawan Tamnar⁴

¹Department of Mechanical Engineering, PREC Loni, India

²Department of Mechanical Engineering, PREC Loni, India

³Department of Mechanical Engineering, PREC Loni, India

⁴Department of Mechanical Engineering, PREC Loni, India

ABSTRACT

Sign language is the only tool of communication for the person who is not able to speak and hear anything. Sign language is a boon for the physically challenged people to express their thoughts and emotion. In this work, a novel scheme of sign language recognition has been proposed for identifying the alphabets and gestures in sign language. With the help of computer vision and neural networks we can detect the signs and give the respective text output.

Keywords: Sign Language recognition, convolutional Neural Network, Image Processing, Edge Detection, Hand Gesture Recognition

1. INTRODUCTION

Speech impaired people use hand signs and gestures to communicate. Normal People face difficulty in understanding their language. Hence there is a need of a system which recognizes the different signs, gestures and conveys the information to the normal people. It bridges the gap between physically challenged people and normal people.

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps:

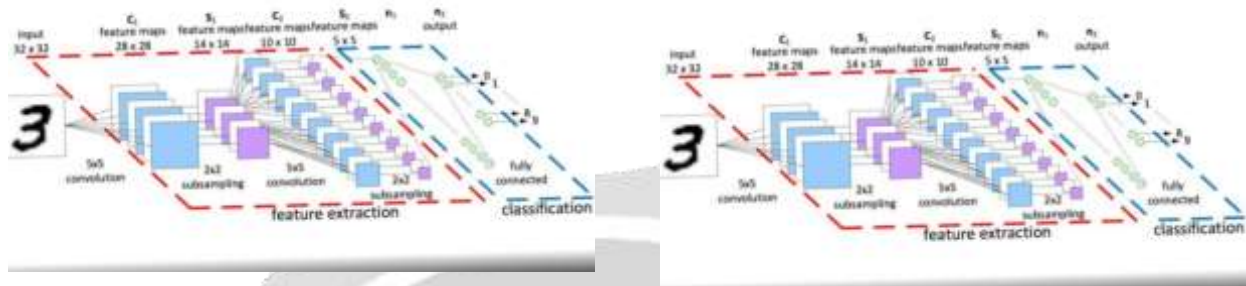
- 1.Importing the image via image acquisition tools.
- 2.Analysing and manipulating the image.
- 3.Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

Digital image processing consists of the manipulation of images using digital computers. Its use has been increasing exponentially in the last decades. Its applications range from medicine to entertainment, passing by geological processing and remote sensing. Multimedia systems, one of the pillars of the modern information society, rely heavily on digital image processing. Digital image processing consists of the manipulation of those finite precision numbers. The processing of digital images can be divided into several classes: image enhancement, image restoration, image analysis, and image compression

1.1 CNN Algorithm

A convolution neural network (CNN) is a type of artificial neural network that is mostly used for image recognition and processing. Although a CNN is a powerful tool, its training process necessitates millions of labelled data points. To train CNN, high-power processors like a GPU or an NPU must be used. This will enable CNN to provide findings quickly enough to be useful. While CNN are intended to handle imagery-related problems, they also have a wide range of uses that go beyond image analysis and recognition, such as image categorization, natural language processing, drug development.



2 LITERATURE REVIEW AND OBJECTIVE

2.1 Objective

A Hand Gesture Speaking System goal is to create a complete system to detect, recognise, and interpret the hand gestures through computer vision.

2.To provide a new low cost, high speed and color image acquisition system.

3.To generate a complete system which can identify, spot and explain the hand motioning through computer sight.

2.2 Literature Review

Many machine learning algorithms can be used to construct handwritten digit recognition systems. The amount of time needed to train the model and its final accuracy can change when different methods are used to this particular task. There is typically a trade-off between time and accuracy when building such systems using different methods. Compared to other algorithms, some techniques take less time to train the model completely, but as a result, they offer less accuracy. On the other hand, many alternative methods require a lot of time to fully train the model yet ultimately produce results that are more accurate. Because to the variety of applications for this technology, some of them demand greater accuracy to function properly and have no time constraints, whilst other applications would need the model.

These algorithms can be compared in light of the fact that the outputs of these algorithms differ in terms of timing and accuracy. Our research focuses more on accuracy, and we'll be evaluating several algorithms on it. In our suggested project, Hand Gesture Speaking System system has been implemented using the advantages of a convolutional neural network. Our model was trained using a number of convolution and pooling layers to achieve high accuracy. Our research revealed that CNN is the most accurate algorithm that can be used to achieve maximum accuracy.

3 MATERIALS AND METHODS

Most of the researchers classified gesture recognition system into mainly three steps after acquiring the input image from camera(s), videos or even data glove instrumented device.

3.1. Extraction Method and image pre-processing Segmentation process is the first process for recognizing hand gestures. It is the process of dividing the input image (in this case hand gesture image) into regions separated by boundaries [12]. The segmentation process depends on the type of gesture, if it is dynamic gesture then the hand gesture need to be located and tracked [12], if it is static gesture (posture) the input image have to be segmented only. The hand should be located firstly, generally a bounding box is used to specify the depending on the skin color [13] and secondly, the hand have to be tracked, for tracking the hand there are two main approaches; either the video is divided into frames and each frame have to be processed alone, in this case

the hand frame is treated as a posture and segmented [12], or using some tracking information such as shape, skin color using some tools such as Kalman filter[12]

The common helpful cue used for segmenting the hand is the skin color [12], since it is easy and invariant to scale, translation, and rotation changes [14]. Different tools and methods used skin and non-skin pixels to model the hand. These methods are parametric and non-parametric techniques, Gaussian Model (GM) and Gaussian Mixture Model (GMM) are parametric techniques, and histogram based techniques are non-parametric. However it is affected with illumination condition changes abs different races [6]. Some researches overcome this problem using data glove and colored markers which provide exact information about the orientation and position of palm and fingers [12]. Others used infrared camera [6], and range information generated by special camera Time-of-Flight (ToF) camera [12], although these systems can detect different skin colors under cluttered background but it is affected with changing in temperature degrees besides their expensive cost [6]. The segmentation considered as an open issue problem itself [9]. The color space used in a specific application plays an essential role in the success of segmentation process, however color spaces are sensitive to lighting changes, for this reason, researches tend to use chrominance components only and neglect the luminance components such as r-g, and HS color spaces. However there are some factors that obstacle the segmentation process which is [12]; complex background, illumination changes, low video quality

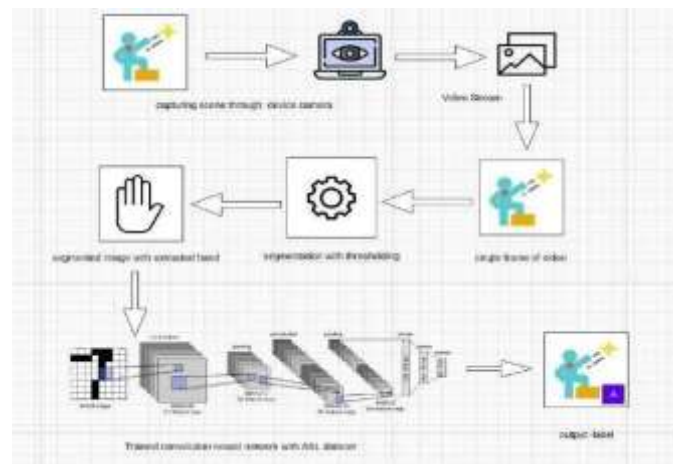
3.2 Feature Extraction

Good segmentation process leads to perfect features extraction process and the latter play an important role in a successful recognition process [6]. Features vector of the segmented image can be extracted in different ways according to particular application. Various methods have been applied for representing the features can be extracted. Some methods used the shape of the hand such as hand contour and silhouette [6] while others utilized fingertips position, palm center, etc. [6] created 13 parameters as a feature vector, the first parameters represents the ratio aspect of the bounding box of the hand and the rest 12 parameters are mean values of brightness pixels in the image. [14] used Self-Growing and Self-Organized Neural Gas (SGONG) neural algorithm to capture the shape of the hand, then three features are obtained; Palm region, Palm center, and Hand slope. [16] calculated the Center Of Gravity (COG) of the segmented hand and the distance from the COG to the farthest point in the fingers, and extracted one binary signal (1D) to estimate the number of fingers in the hand region. [15] divided the segmented image into different blocks size and each block represents the brightness measurements in the image. Many experiments were applied to decide the right block size that can achieve good recognition rate [15]. [17][18] used Gaussian pdf to extract geometric central moment as local and global features

3.3 Gesture Classification

After modeling and analysis of the input hand image, gesture classification method is used to recognize the gesture. Recognition process affected with the proper selection of features parameters and suitable classification algorithm [7]. For example edge detection or contour operators [9] cannot be used for gesture recognition since many hand postures are generated and could produce misclassification [9]. Euclidean distance metric used to classify the gestures [19][5][17]. Statistical tools used for gesture classification, HMM tool has shown its ability to recognize dynamic gestures [20][13]besides, Finite State Machine (FSM) [21], Learning Vector Quantization [22], and Principal Component Analysis (PCA) [23]. Neural network has been widely applied in the field of extracted the hand shape [14], and for hand gesture recognition [24][25][26]. Other soft computing tools are effective in this field as well, such as Fuzzy C-model, 60,000 28*28 grayscale photos are used. Our model is trained through a conventional 5 epochs to reach accuracy of the order of 99.16% which is substantially higher as compared to the traditional techniques such as SVM, Multi layer Perceptron, Bayes Net, Random Forest, etc. used to create handwritten digit recognition systems. Means clustering (FCM) [6], and Genetic Algorithms GAs [27]. Figure 4 explain the architecture of classification system.

3.3 Proposed System



A. Image Processing

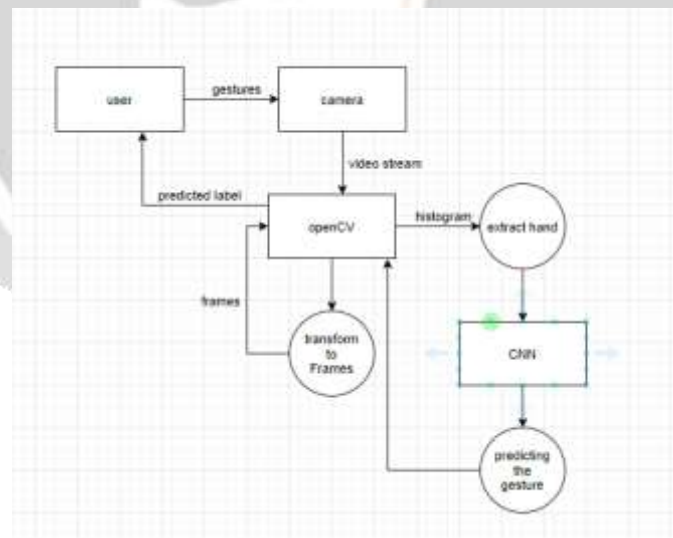
Image processing is the process of implementing various methods such as re sizing the images, converting them to grayscale format, and augmentation of images for enabling the digital image data to be used within the machine learning model effectively.

B. Training Neural Network

After completing data processing, the CNN model will be created which consists of various convolution and pooling layers alongside a 3x3 sized kernel. The model will then be trained on the basis of training and validation data with the help of several python libraries such as TensorFlow, Pillow, Open CV, that were preloaded to perform these specific tasks.

C. Testing Accuracy of Neural Network

After the model is trained using the training dataset, we use the testing dataset to evaluate how well it works.



CONCLUSIONS

The importance of gesture recognition lies in building efficient human-machine interaction. This paper describes how the implementation of the system is done based upon the images captured. Hand detection is done using OpenCV and TensorFlow object detector. And further it is enhanced for interpretation of gestures by the computer to perform actions like switching the pages, scrolling up or down the page.

ACKNOWLEDGEMENTS

I thank Mr..S.R Nimbalkar, Prof of B.E Mechanical, PREC, Loni, for his continuous support and encouragement for completing this research paper and also thanks to PREC Mechanical department for the continuous support.

REFERENCES

- [1] G. R. S. Murthy, R. S. Jadon. (2009). "A Review of Vision Based Hand Gestures Recognition," International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405-410.
- [2] P. Garg, N. Aggarwal and S. Sofat. (2009). "Vision Based Hand Gesture Recognition," World Academy of Science, Engineering and Technology, Vol. 49, pp. 972-977.
- [3] FakhreddineKarray, MiladAlemzadeh, Jamil AbouSaleh, Mo Nours Arab, (2008). "Human-Computer Interaction: Overview on State of the Art", International Journal on Smart Sensing and Intelligent Systems, Vol. 1(1).
- [4] Wikipedia Website.
- [5] Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 3(2).

