International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 12 Issue 6, June, 2023, Impact Factor: 7.751

Real Time Hand Gesture Recognition

Dikshitha Kamath K[1], Dr. Bharathi Pilar[2]

[1]Research Scholar, Mangalore University, Konaje
[2]Assistant Professor Department of Computer Science University College Mangalore

ABSTRACT

Hand gestures can play important role in Computer Vision as well as communication through sign language. It provides an interaction between human and machine. Deafness is a degree of loss where the person could not understand speech and spoken language. Sign language faces many difficulties through hearing. Real time hand gesture recognition has been proposed in our research. Our proposed CNN model is used to communicate with deaf people. The Proposed model has achieved an accuracy of 94.6% to recognize different gestures.

General Terms: Hearing impaired people, Computer Vision, Hand gesture recognition.

Keywords: Sign Recognition, Gesture Recognition, Convolutional Neural Networks.

INTRODUCTION

Disability can be considered as an absolutely necessary subject concerned with human rights. Disability can be classified into different types: Hearing disability, Visual disability, Physical disability, Speech disability and mental disability. Hearing is one of the necessary senses in every individual. The person who is incapable of hearing that is hearable to others is Hearing Loss Person. Hearing Loss leads to severe problem. More than 6% of the World people – approximately 468 million people have the hearing problem (433 million adults and 35 million children) [1]. Deaf and Hearing impaired people can apply sign language which is helpful for communication.

Sign Language is used to represent thinking which includes different body or Arms Postures which is a combination of several hand position and shape, Facial expressions. Sign Language is not a universal language. However, there exist variety of Sign Languages vary from country to country and culture to culture. Sign Languages classified into American Sign Language, Japanese Sign Language, Indian Sign Language, Arabic Sign Language and so for respective countries also. People can't assume a single instant without communication. Different types of technical process supporting the communication. But few physically disabled people do not get the advance support for any technical improvements. As a result it is not easy to communicate with sign language users with normal people because of insufficient attention to communicate with deaf and dumb people. Loss of hearing has a good meaningful impression on a child cultural improvements and education achievements. Deaf learning focused in many developed countries but the development needed in developing countries as Bangladesh. Deaf and dumb people can be considered an unreached people group. Because lack of importance in society. As a result, sign language users cannot enjoy services like education, health services and employment opportunities. Huge amount of research project applied to hand gesture recognition using excellent technological support. This framework is designed to communicate between deaf people and normal people.

RELATED WORKS AND BACKGROUND

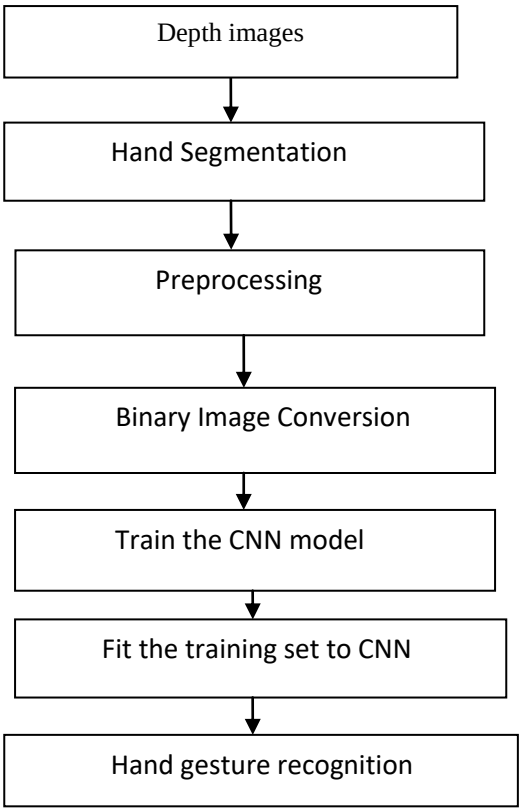
The Hand gesture is one medium of signing. Gesture recognition has become one of the most interesting topics in Computer vision and Deep Learning. Firstly the glove based system was developed to recognize hand gestures, but the interface restricts the user comfort because of wearing heavy devices and the burden of cable communication [2].Though it provide low cost hardware portability facility and charge transfer naturalness [3]. For reaching out to the problem of impaired dedicated to translating ASL (American Sign Language). 3D depth sensing camera was used image acquisition together with applying depth thresh holding, contour, convex hull and convexity defects were applied to the binary image for Sign Language classification [5].Shukla et.al used the Naïve Bayes classification algorithm to classify five fingers only. In [6] the color skin information is used to segment the hands for the generic hand based

remote control system. Commonly hand gesture recognition system follows three phases: detection, tracking, and recognition. Survey [4] provides perceptron of existing methods of gesture recognition systems for Human Computer Interaction by categorizing different key parameters.

In a couple of years deep learning approaches have shown a flawless achievements [7] to suppress traditional Machine learning techniques in computer vision. Peijun Bao et al [7] proposed Deep CNN to directly recognize the gesture from the whole image without using any region proposal algorithm or sliding window mechanism. On the other hand, recognition of gestures with both written words and audible speech was proposed [8] by image matching technique and use of build in "Find" function. Preceding analysis on different sign languages such as American, Indian [9] and Arabic as well as Italian has fascinated researchers to work with Bangla Sign Language. Rahman et al proposed Haar-Like feature based cascade; Here and saturation implemented for hand sign extraction and K-nearest Neighbour (KNN) algorithm for 36 Bangla character recognition yielded an average accuracy of 96.46% [10]. A Real time sign Language has been proposed for the ILSVRC 2012 dataset for American Sign Language using Google Net architecture [11]. As human activity and gesture recognition has become one of the growing domains of ambient intelligence, The author proposed Convolutional neural network (CNN) and recurrent neural network (RNN) for automated hand gesture recognition using both depth and skeleton data. This method achieved overall accuracy of 85.46% for 14 dynamic gestures [12]. As main focus on depth images, So in [15] using senz3d dataset depth images, real time hand gesture recognition was proposed by Memo et al. Hand was segmented by hand contour and multi class support vector machine was used to classify 11 gestures that gained accuracy 90%[15]. This research objective is to facilitate the learning and communication process of deaf and hearing impaired people.

PROPOSED METHODOLOGY

In our research, the hand gesture recognition system has been proposed by hand segmentation and preprocessing operation followed by the CNN Model architecture. However this image classification models become prominent in computer vision field. Here after acquiring the depth images segment the hand data which follows some preprocessing activities train the CNN model for feature extraction and gesture recognition. In the figure Block diagram of proposed methodology has been shown below:



Proposed Methodology of the framework

Hand segmentation

In our work, depth images have taken from the dataset. In Hand gesture recognition the foremost task is extracting the hand regions. So our first step is to segment the hand region from the depth map. This implies that hand is close to camera and begins with thresh holding of the depth values. After acquiring the depth images color space has been implemented to segment the hand.

Preprocessing

It is challenging where how to prepare image data when training a convolutional neural network [19]. Images in training dataset had different sizes. It is necessary to resize it being used as input to the model [19]. This consists of both resizing and cropping techniques during both training and evaluation of the model [19].

Binary Image Conversion

It is easier to deal with binary image than multiple color channels. Thresh holding convert images to binary images. Pixel below threshold value (60) converted to 0 and above threshold value converted to 1.

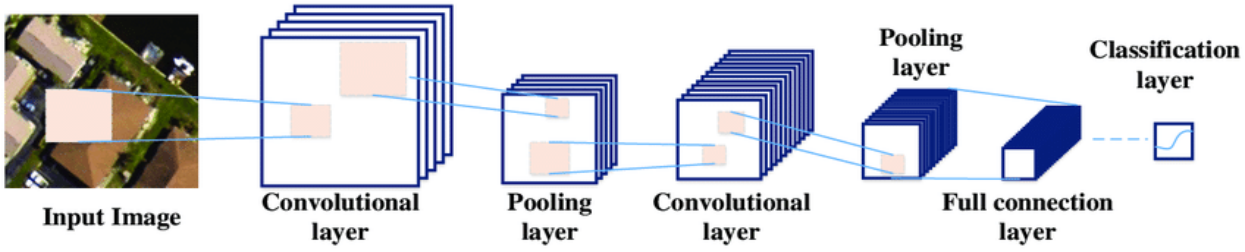
EXPERIMENTAL SET UP

The regular version of multilayered perceptron is CNN. It works on neurons. The idea of neuron based on the working of the neurons of the animal visual cortex. The idea is that image is converted into 2 D array of numbers and given to the computer. The work is similar to how human brain works. When we look at a picture of a dog, we can classify it as such if the picture has identifiable pictures such as paws or four legs. In a similar manner, computer performs image classification for low level features such as edges and curves through a series of convolutional layers. The Computer uses low level features obtained at the initial level to generate high level features such as eyes to identify the object.

Layers of CNN

A Convolutional neural network consists of three layers that is an Input layer, Hidden layers and an output layer. In CNN the hidden layer performs convolution. Typically, the layer does multiplication or dot product and its activation function is commonly RELU layer followed by other convolution layers Pooling layers, Fully connected layers, Normalization layer and convolutional layers.

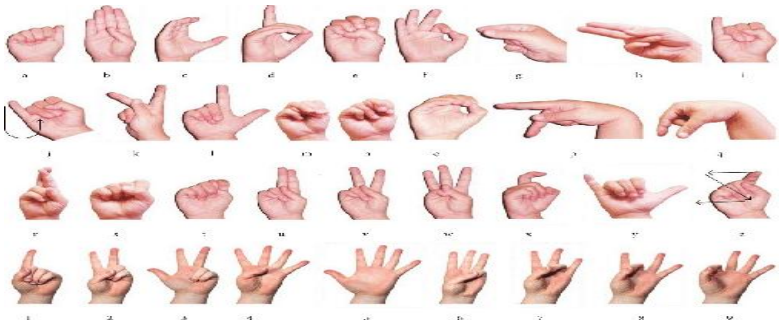
A Convolutional layer will take the input and passes to the next layer. It depends on following attributes kernel size, Number of input channels and output channel, Depth of Convolutional filter must be equal to the number of channels. Pooling layer decrease the data dimension. It combines the output of neuron at one layer into single neuron of the next layer. It is of two types Max Pooling and average Pooling uses the average value through the cluster. Fully connected layer is the final stage of CNN. Every neuron of one layer is connected to every neuron of another layer. The flattened matrix goes to fully connected layer for classification.



Proposed CNN Model

Dataset Description

A dataset is necessary to gain more immeasurable quality. It is the most crucial aspect that makes algorithm training possible. Here senz3d dataset of hand gestures has been used [13] [14] [15]. To work with depth images dataset includes various different static gestures obtained by creative senz3d camera [13] [14]. The dataset consists of 1320 sample images composed of four different people images. Each person made different gestures repeatedly for 30 times. Each sample, color, depth and confidence frame available for each dataset. Snapshot of few images of the dataset for hand gesture recognition given in the figure.



Sample Images of the Dataset

Evaluation Metric

Sample index: The test data of the algorithm depends on number of samples which decides the performance of recognition. The recognition is successful only if it constitutes more number of samples with variations shows the algorithm generality and robustness.

Complexity Index: The computational complexity of gesture recognition depends on the number of classes which decides its success in real world implementation.

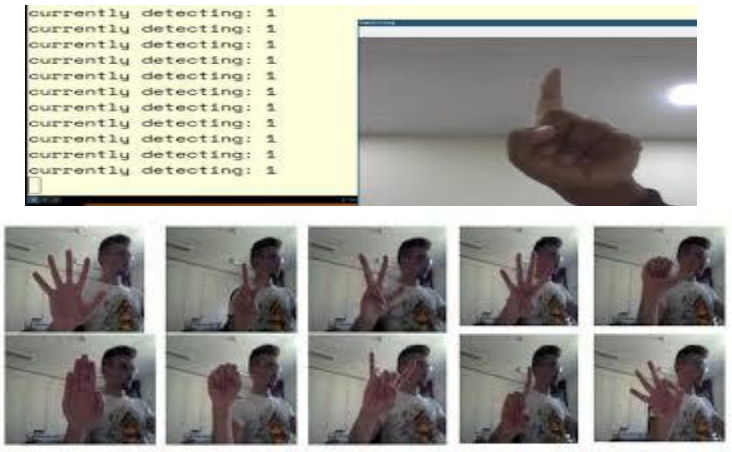
Dataset Index: Sharing data and code is important for replication of system and community. Dataset can be published which will help other researchers to verify the results.

Availability Index: Algorithm should make available helps other researchers to recreate study and evaluate results objectively.

Background Index: Background of real visual scenes is complex. The developer may consider complex and cluttered backgrounds with gesture patterns to be recognized.

RESULT ANALYSIS

In this paper we discussed a method of recognizing hand gestures using CNN. The main work here is to identify the gesture provided by the user. Dataset plays a significant role in image classification and identification. The variation in the dataset and number of images is more then more is the accuracy. Since large dataset requires huge amount of time to compile, a PC with high graphics and RAM can solve this problem. The dataset contain 600 images of each hand sign. Since PC not so powerful to handle large dataset, It will take much time even if somehow compiles the program. It can also be noticed that camera plays a important role in image classification. When image quality is not clear it is difficult to recognize images. Since low quality image have low pixel, therefore they give more information about the image. Here Accuracy has converged near to 94.6% for different gestures.



Output of real time hand gesture recognition using proposed methodology

Class	Accuracy
0	0.9375
1	0.9374
2	0.9642
3	0.9774
4	0.9841
5	0.9542
6	0.9458
7	0.9462
8	0.9647
9	0.9420
10	0.9841

Performance of the proposed architecture hand gesture recognition

CONCLUSION

In this paper we used CNN to recognize hand gestures and discussed various parameters depending on accuracy. Open CV used to take real time video and recognize the gestures. The probability of gestures is arranged in descending order. We have seen that the accuracy is 94.6% for different gestures from depth images depends on the number of factors like camera, dataset, and techniques being used.

REFERENCES

[1]. W.H. Organization et al. 2020 Deafness and hearing loss fact sheet 300. Updated on March 2020.
[2]. Pavlovic VI, Sharma, R and Huang, TS,1997 visual interpretation of hand gestures for human computer interaction : A review . IEEE Transactions on pattern analysis and machine intelligence.
[3]. Abhishek,KS; Qubeley, LCF and Ho,D; 2016, August Glove based hand gesture recognition sign language translator using capacitive touch sensor. In 2016 IEEE International conference on Electronic devices and solid state circuits IEEE.
[4]. Rautaray, S.S and Agrawal, A, 2015. Vision based hand gesture recognition for human computer interaction : a survey .Artificial Intelligence review.
[5]. Shukla J and Dwivedi A,2014 April. A method for hand gesture recognition. In 2014 Fourth international conference on communication systems and network technologies.
[6]. Erden, F and cetin, A.E; 2014. Hand gesture based remote control system using infrared sensors and a camera .IEEE Transactions on Consumer Electronics.
[7]. Bao, P; Maqueda, AI; del-Blanco, CR and Gracia N 2017. Tiny hand gesture recognition without localization via a deep convolutional network. IEEE Transactions on Consumer Electronics
[8]. Akoum,A and AI Mawla, N,2015. Hand gesture recognition approach for asl language using hand extraction algorithm. Journal for software engineering and applications.
[9]. Akoum, A and AI Mawla, N; 2015. Hand gesture recognition approach for asl language using hand extraction algorithm. Journal of software engineering and applications.
[10]. Rahaman MA; Jasim, M; Ali, M.H and Hasanuzzaman M; 2014 December real time computer vision based Bengali sign language recognition. In 2014 17th International conference on computer and Information technology. IEEE
[11]. Garcia, B and viesca, SA; 2016. Real time American sign language recognition with covolutional neural network for visual recognition.
[12]. Lai, K and yanushkevich, S.N; 2018, August. CNN+RNN depth and skeleton based dynamic hand gesture recognition. In 2018 24th international conference on Pattern recognition. IEEE.
[13]. G. Marin, F Domino, P Zanuttigh, "Hand gesture recognition with leap motion and Kinect devices" IEEE International conference on Image Processing, Paris, France, 2014.
[14]. G Marin, F. Domino, P. Zanuttigh, " Hand gesture recognition with jointly calibrated leap motion and depth sensor", Multimedia tools and applications, 2015.
[15]. Minto L and Zanuttigh, P, 2015. Exploiting silhouette descriptors and synthetic data for hand gesture recognition.

[16]. Santosh, K.C and hegadi , R.S.eds,2019. Recent trends in Image processing and pattern recognition.Second international conference, RTIP2R 2018 Solapur, India, December2018.
[17]. Hossain, S Sarma, D Mitra, T alam, M.N SAHA, I and Johora, FT 2020, July , Bengali Hand sign gesture recognition using Convolutional neural network.
[18]. Islam, MS, Rahman, MH , Arifuzzaman, m, Sassi , R and Aktaruzzaman, M ; 2019,September. Recognition bangla sign language using convolutional neural network.In 2019 International conference on innovation and Intelligence for informatics, Computing, and technologies.
[19]. Brownlee,J;2019. Deep learning for computer vision: Image classification, object detection and Face recognition in python. Machine Learning Mastery.
[20]. Saha, S; 2018. A comprehensive guide to convolutional neural network – The EL15 way. Towards data science.
[21]. Brownlee, j; 2019. A gentle introduction to the rectified linear unit. Machine learning mastery.