Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 3, March 2019, pg.352 – 357

Recognition of American Sign Language using Image Processing and Machine Learning

Sharmila Gaikwad[1]; Akanksha Shetty[2]; Akshaya Satam[3]; Mihir Rathod[4]; Pooja Shah[5]

[1]Department of Computer Engineering, MCT's Rajiv Gandhi Institute Of Technology, Mumbai, India
[2]Department of Computer Engineering, MCT's Rajiv Gandhi Institute Of Technology, Mumbai, India
[3]Department of Computer Engineering, MCT's Rajiv Gandhi Institute Of Technology, Mumbai, India
[4]Department of Computer Engineering, MCT's Rajiv Gandhi Institute Of Technology, Mumbai, India
[5]Department of Computer Engineering, MCT's Rajiv Gandhi Institute Of Technology, Mumbai, India
[1] sharmila.gaikwad@mctrgit.ac.in; [2] shettyakanksha97@gmail.com; [3] akshayasatam@gmail.com,
[4] mihirrathod2723@gmail.com, [5] pooja.shah1247@gmail.com

Abstract— In this era of touch screen devices, human machine interaction is a very important factor nowadays. Many devices are developed which can be operated without touching the system. So, in this project we are going to demonstrate how one can operate the system without actually touching the system but with the use of gestures. Human–Computer Interaction (HCI) can be described as the point of communication between the human user and a computer. Commonly used input devices include the following: keyboard, computer mouse, trackball, touchpad and a touch-screen. All these devices require manual control and cannot be used by persons impaired in movement capacity. Therefore, there is a necessity for developing various strategies of communication between human and personal computers that would be required for the persons with motor impairments and would enable them to become a part of the Information Society. In this project we will be implementing a system which will include Hand gesture recognition approach for ASL language - Sign language being a method used for communicational purposes by deaf people. We will discuss the different steps used to provide the system with the required information, to identify the provided data and further analyse the hand gestures making use of various algorithms. Hence, this system will give them a new way to interact with the computer world and the objective of connecting the specially abled with the computer world will be sufficed with the help of this project.

Keywords— Human Computer interaction, hand gesture recognition, Sign Language

I. INTRODUCTION

Sign language is the most expressive way for communication between hearing impaired people, where information is majorly conveyed through the hand/arm gestures. SLR plays an important predominant role in developing the gesture-based human–computer interaction systems. The recent years of statistics have witnessed an increased research interest in interaction and intelligent computing. In the present-day framework, computers have become a key element of our society for interaction. As the hand signs constitute a powerful interaction human communication modality, they can be considered as an intuitive and convenient mode for the communication between normal human and human with hearing problems. Several Hand gesture recognition methods and already subsist and most of them are based on Hidden Markov Models, Fuzzy Logic, Neural Networks, etc [1], [2], [3]. Even though most of the methods are extremely effective the computation cost of these methods are high. In order to overcome this, we have developed an alternative method to recognize American sign language with the help of hand gestures.

II. Related Works

Convolutional Neural Networks have been extremely successful in image recognition and classification problems, and have been successfully implemented for human gesture recognition in recent years. In particular, there has been work done in the realm of sign language recognition using deep CNNs, with input-recognition that is sensitive to more than just pixels of the images. With the use of cameras that sense depth and contour, the process is made much easier via developing characteristic depth and motion profiles for each sign language gesture. The use of depth-sensing technology is quickly growing in popularity, and other tools have been incorporated into the process that have proven successful.

Developments such as custom-designed color gloves have been used to facilitate the recognition process and make the feature extraction step more efficient by making certain gestural units easier to identify and classify. Until recently, however, methods of automatic sign language recognition weren’t able to make use of the depth-sensing technology that is as widely available today. Previous works made use of very basic camera technology to generate datasets of simply images, with no depth or contour information available, just the pixels present. Attempts at using CNNs to handle the task of classifying images of ASL letter gestures have had some success, but using a pre-trained GoogleNet architecture.

III.Literature Survey

The depth sensors, like the Xtion PRO LIVE sensor, have given rise to new opportunities for human-computer interaction (HCI). Although great progress has been made by using the Xtion PRO LIVE sensor in human body tracking and body gesture recognition, robust hand gesture recognition still remains a problem. Compared to the human body, the hand is a smaller object and has more complex articulations. Thus, a hand is easily affected by segmentation errors as compared to the entire human body.

C.W.Ng and S.Ranganath interpret a user’s gestures in real-time using hand segmentation to extract binary hand blobs. The shape of blobs is represented using Fourier descriptors. This Fourier descriptor representation are input to radial basis function(RBF) networks for posture classification.

N.Tanibata et al. obtain hand features from a sequence of images. This is done by segmenting and tracking the face and hands using skin colour. The tracking of elbows is done by matching the template of an elbow shape. The hand features like area of hand, direction of hand motion, etc. are extracted and are then input to Hidden Markov Model (HMM).

D.Kelly et al. recognise hand postures used in various sign languages using a novel hand posture feature, eigen-space Size Function and Support Vector Machine (SVM) based gesture recognition framework. They used a combination of Hu moments and eigen-space Size Function to classify different hand postures.

H. K. Nishihara et al. (US patent, 2009), generate silhouette images and three-dimensional features of bare hand. Further, classify the input gesture by comparing it with predefined gestures. Daniel Martinez Capilla used 8dimensional descriptor for every frame captured by Microsoft Kinect XBOX 360 and compared the signs by dynamic time-warping (DTW).

Jagdish L. Raheja et al. This paper describes the use of fingertips, centre of palms detection in dynamic hand gestures produced by either one or both hands without utilizing any sensor or marker. Natural Computation as no sensor, colour is used on hands for segmentation of the picture and hence would enable the user to perform operation using their bare hands. Yi Li This system consists of three components: Hand Detection, Finger Identification, and Gesture Recognition. The system is built on Candescent NUI project, which is freely available online. Open NI framework was used to extract depth data from the 3D sensor.

IV.System Flow

Our paper was inspired by the work done by Alhussain Akoum and Nour Al Mawla [7], where the primary focus is the ASL system wherein with the use of various data extraction and data matching algorithms effective results were obtained. The system proposed in our case accepts video as an input. Our system not only identifies the American sign language but also provides the respective user with a text to speech conversion.

A. Methodology

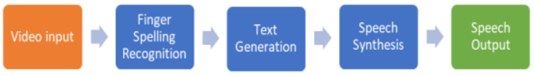

Fig 1: Process Block Diagram

a) Image Acquisition

The initial phase of Image Acquisition is obtaining a picture during runtime via integrated camera and while obtaining these pictures will be loaded in the directory after they are captured. The recently captured image will be obtained and will be compared to the images stored for specific letter in the database making use of the SIFT algorithm. The comparison will give the gesture that was done as an output and also the translated text for the following gesture. The images will be captured via the code of opening a webcam through OpenCV and frames will be captured every second which will be stored in another directory where all the inputs images are stored in another directory and then comparison of the captures image and the pre-stored images are made.

b) Feature Extraction

For any of the object, there are many features, interesting points on the object, which can be extracted to provide a "feature" description of the object. SIFT picture features produces a certain set of features which are not affected by many of the complications experienced in object scaling and rotation. SIFT approach, for generation of the image feature, takes a picture and transform it into a "big collection of local feature vectors". Each of the feature vectors never changes to any of scaling, rotation or translation of the image.

c) Orientation Detection

This step will take the input of hand movement in any of the form and the gesture will be identified by the described part of feature extraction since the SIFT algorithm includes the orientation assignment procedure.

d) Gesture Recognition

On the completion of the entire process the application will be then translated into its recognized character or alphabet from the gesture which is beneficial to be understood in layman's language. The following process contains passing out the 1- dimensional array of 26 characters corresponding to alphabets has been passed where the image number stored in the database is provided in the array.

e) Text to Speech

When the character is get selected based on recognized sign using speech conversion, recognized text is converted to speech and an audio output is executed.

B. Implementation

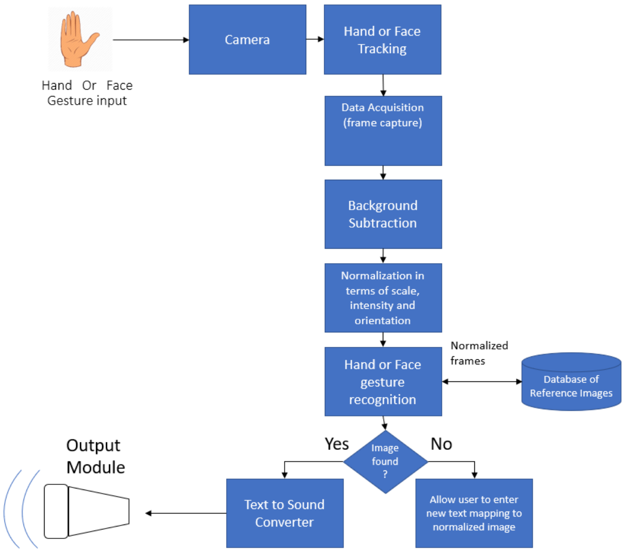
Hand Or Face
Gesture input
Camera
Hand or Face Tracking
Data Acquisition (frame capture)
Background Subtraction
Normalization in terms of scale, intensity and orientation
Normalized frames
Database of Reference Images
Hand or Face gesture recognition
Yes
Image found ?
No
Output Module
Text to Sound Converter
Allow user to enter new text mapping to normalized image

Fig 2: System Flow

a) Image Acquisition Model:

Image acquisition is the process of creating photographic images, such as the interior structure of an object. The term is often assumed to include the compression, storage, printing, and display of such images.

b) Pre-processing Model:

The main aim of pre-processing is an improvement of the image data that reduce unwanted deviation or enhances image features for further processing. Pre-processing is also referred to as an attempt to capture the important pattern which expresses the uniqueness in data without noise or unwanted data which includes cropping, resizing and grey scaling.

c) Cropping:

Cropping refers to the removal of the unwanted parts of an image to improve framing, accentuate subject matter or change aspect ratio.

d) Resizing:

Images are resized to suit the space allocated or available. Resizing image are tips for keeping the quality of the original image. Changing the physical size affects the physical size but not the resolution.

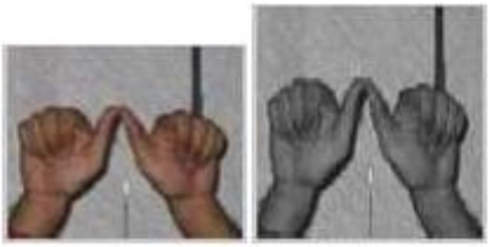

Fig 3: Output of Pre-processing

e) Feature Learning:

Comprised of one or more convolutional layers and followed by one or more fully connected layers as in a standard multilayer neural network. It implicitly extracts relevant features from a Feed-forward network that can extract topological properties from an image. Like almost every other International Journal on Recent and Innovation Trends in neural networks, CNNs are trained with a version of the backpropagation algorithm.

f) Convolutional layer:

Core building block of a CNN. Layer's parameters consist of a set of learnable filters (or kernels). The filter is convolved across the width and height of the input volume. Computing the dot product between the entries of the filter

g) Pooling Layer:

Reduce the spatial size of the representation to reduce the number of parameters. Independently operates on every depth slice of the input. The most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples every depth slice in the input by 2 along with both the width and the height, discarding 75% of the activations spatially, using the MAX operation.

h) ReLU layer:

ReLU is the abbreviation of Rectified Linear Units which increases the nonlinear properties

i) Fully connected layer:

Neurons in a fully connected layer have full connections to all activations in the previous layer. The activations are computed with the matrix multiplication.

V. PROPOSED ARCHITECTURE

Most implementations surrounding this task have attempted it via transfer learning, but our network was trained from scratch. Our general architecture was a fairly common CNN architecture, consisting of multiple convolutional and dense layers. The architecture included 3 groups of 2 convolutional layers followed by a max-pool layer and a dropout layer, and two groups of fully connected layer followed by a dropout layer and one final output layer.

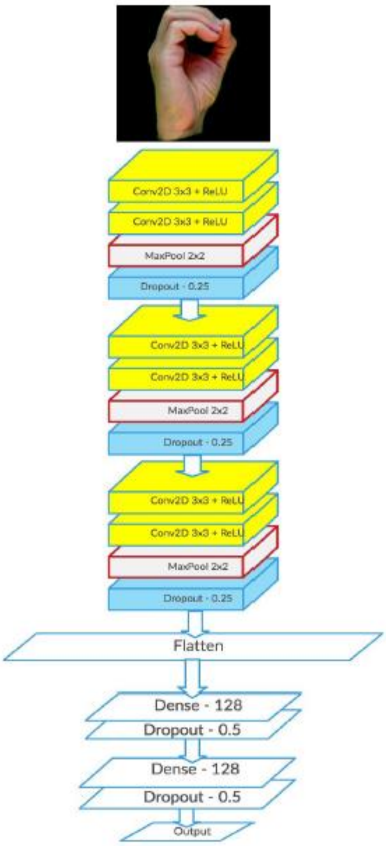

Fig 4:Layered Architecture

VI. CONCLUSIONS

Nowadays, hand gesture recognition had been implemented into many forms of bids and in several. This is a proof of the importance and improvement of this research title over the past few years. Hand gestures are at the heart of vision collaboration and machinery control. This paper principally dedicated to the ASL system. The system will provide an interface that can easily communicate with deaf people by Sign Language Recognition. The system is not only can apply in family environment, but also can apply in public. For the Social use this system is very helpful for deaf and dumb people. We will build simple gesture recognizer based on OpenCV toolkit and integrated it into Visionary framework. The project involves distinguishing among different alphabets of English language. Also, the project is included with recognition of all the English alphabets & numbers. Further, we may move on to recognizing words, from as large a dictionary as possible, from Indian Sign Language.

ACKNOWLEDGEMENT

We wish to express our sincere gratitude to the management, Dr. Sanjay Bokade, Principal and Dr. Satish Ket, H.O.D of Computer Department of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our technical paper on Gesture Recognition for Specially Abled.

REFERENCES

[1] Supawadee Saengsri, Vit Niennattrakul, and Chotirat Ann Ratanamahatana. 'TFRS: Thai Finger-Sign Sing Language Recognition System', IEEE, pp: 457- 462, 2012.

[2] Subha Rajam, P. and Balakrishnan, G.,'Recognition of Tamil Sing Language Alphabet using Image Processing to aid Deaf-Dumb People', International Conference on Communication Technology and System Design, Elsevier, pp: 861-868, 2011.

[3] Subha Rajam, P. and Balakrishnan, G., 'Real Time Sign Language Recognition System to aid Deaf-dumb People', IEEE, pp: 737- 742, 2011.

[4] Chance, M. Glenn, Divya Mandloi, Kanthi Sarella, and Muhammed Lonon, (2005) 'An Image Processing Technique for the Translation of ASL Finger-Sign to Digital Audi or text', NTID International Instructional Technology and Education of the Deaf Symposium, pp: 1-7.

[5] Nicolas Pugeault, and Richard Bowden,'Sign It Out: Real-Time ASL Fingersign Recognition', IEEE Workshop on Consumer Depth Cameras for Computer Vision, pp. 1-6, 2011.

[6] A. S. Ghotkar, R. Khatal, S. Khupase, S. Asati, and M. Hadap, "Hand Gesture Recognition for Indian Sign Language", IEEE International Conference on Computer Communication and Informatics (ICCCI), Jan. 10-12, 2012, Coimbatore, India.

[7] Alhussain Akoum and Nour Al Mawla 'Hand Gesture Recognition Approach for ASL Language Using Hand Extraction Algorithm', Journal of Software Engineering and Applications, 2015, 8,419-430.