System For Improved Sign Language Recognition

Abstract—

Sign language is one of the common methods of communication among vocally and audibly impaired people. But when it comes to communication between normal and deaf-mute people, this form of communication doesn't help. So far, a whole lot of systems have been developed that interpret sign language. But they often fail to make it effective since they interpret individual words. This provides a poor experience of communication between the people. This paper discusses the development of an interface that takes gestures in sign language as input, and then recognizes the same to provide a voice-based output for corresponding actions in a combinatory manner. The main aspect in such non-verbal communication is the motion of hands, which can be tracked using a 3D depth camera such as a Microsoft Kinect that helps to identify gestures based on skeletal joint movements as well as using RGB data. We propose a real time system that takes individual actions collaboratively and return them as intuitive verbal phrases.

Index Terms - Kinect, Depth Camera, RGB Data, Sign Language.

I. Introduction

Around seven lakh people across the world are vocally impaired. If that be the case, a deaf-muted people will find it difficult to have a conversation with normal people. This gives a poor experience of expressing views and perceptions for both of them. Hence, there is a need of a method that helps a deaf-mute people communicate without a loss in expression. Various approaches were been proposed in order to intake the gestures like Data gloves to sense the data but in intention to give the input gestures by human naturally instead of sticking on electro-magnetic behaviors and also in concerning with high cost later carried out with Histogram of capturing the data with the help of orientation but for the ambiguity of sensing different gestures as the same data it has also had a replacement of representing as like static gestures by following the simple patterns and emerged with dynamic gestures using leap motion controller.

However, for capturing better datasets Skeleton Tracking methodology has been invoked for its high accuracy of perceiving the data which tends to detect the gestures based on skeletal joint movements. Despite its better accuracy, lags behind in detecting variants made in fingers as it targets hands over the fingers. From the reasoning of above standards here, an advanced standard called Kinect X-box 360 is made use of, which aims to intake gestures more accurately through Depth images where depth camera has been incorporated within the sensor. The camera tracks motion by sensing the skeletal joint movements.

II. Existing Method

A.Exploiting Recurrent Neural Networks And Leap Motion Controller For The Recognition Of Sign Language And Semaphoric Hand Gestures – By Danilo Avola and Marco Bernardi

The employed technique Recurrent Neural Network which used to disclose the temporal dynamic behavior in which the input and output states are being controlled by Long-Short term memory. This technique might provide an effective way of representing new datasets based on a large subset. But it does not find a way to identify semaphoric Hand Gestures; hence it fails to extract features with RGB data

B. Static And Dynamic Hand Gesture Recognition In Depth Data Using Dynamic Time Warping- By Guillame Plouffe and Ana-Maria Cretu

The technique deals with representation using Dynamic Time Warping which tends to calculate the optimal match between the sequences and in result could able to recognize both static and dynamic gestures. These impacts on obtaining better detection of background extraction and with skin color. However, absence of a module for a more accurate adaptation of the size of a user hands instead of double training ate different depths.

C. Large Vocabulary Sign Language Recognition Based On Fuzzy Decision Trees-By Gaoling Fang and Wen Gao

The technique involves usage of Fuzzy Decision Trees and Hidden Markov Model which are used to recognize Gesture

Sign Language Recognition

and temporal pattern recognition. This technique comes useful for diminishing classification errors. But a major downside is the inability to blend non-manual parameters into sign language recognition.

D. A Survey On 3d Hand Gesture Recognition- By Hong Cheng and Lu Yang

The paper imposes on sign language interpreter by Hand trajectory and with static gestures. The technique makes it able to recognize gestures using 2D and 3D modeling. Still it needs to have better accuracy in recognition and concerned with irrelevant prediction of gestures.

E. Human Action Recognition By Learning Spatio-Temporal Features With Deep Neural Networks - By Lei Wang and Jun Cheng

The defined method aims to recognize the actions that are being done by the object with Deep Neural Networks through RGB data. This is advantageous since accuracy level in prediction becomes high and fast with the RGB data. But building the neural network is a tedious task

III. Proposed Method

From the consideration of above existed methods in tend to gain recognition Here, we proposed to interpret as verbal form i.e. Sign Language as with the individual wording gestures to recognize by dynamic behavior over just making as actions or activities of what the object does and making the recognition as audible form to get into as real time application. The modules of the task are:

A. Setting up the Kinect

In order to get the gestures to be recognized to intake the input data can be made use through the sensor like Kinect X-box 360. so that to ensure the connectivity between the pc and kinect first as it is used to predict the gestures. To make the connection enable which can be made with the mediatory called Microsoft Kinect where need to install Kinect Studio first (v1.8) and then installing of all necessary drivers that make kinect compatible with the Operation System and then installing Developer Toolkit Browser (v1.8.0) and Skeleton Basics WPF package to sense the skeletal joint movements and finally run SBWPF to start foreground object movement identification.

B. Object Detection

From the making of algorithms to sense the foreground to perceive the input kinect get into advance with the inbuilt functionality of depth camera which is used to detect the object by skeletal joint movements by analyzing the movements of various human joints. The depth camera can be used more effectively to sense both the skeletal joint movements as well as RGB Data so as to make the conversation more intuitive and natural. The proposal is to make the system just like the way human brain interprets sign language. For this purpose, a trained model is used to match across the gestures performed in front of the camera. The basic formation of sign language is just verbs and nouns shown altogether one by without any prepositions and grammar within. So it is enough to have the main words that form the sentence. But since the conversation has to be natural the output should be grammar perfect. Hence here a method that forms sentence in light of the words that are induced by the gestures. Following is the functionality of how kinect works.

1. Segmentation

Segmentation in general refers to separating parts or sections. Here segmentation is done in order to separate the skeletal image from the other objects identified by the kinect. After the image being captured by the kinect, it used to segment out the object which is to be examined for recognition from the background features. This helps in identifying the foreground object with a clear view.

2. Background subtraction

It aims to partition the foreground object from the actual object i.e. hand positions are taken into a part from the human skeleton from the extraction of background features to intake the patterns. The background however is preferred to plain screened so as to carry out an effective background subtraction.

3. Motion detection and thresholding

After examining the foreground object, Thresholding comes into the act by which the thresholding image being converted into Binary image where foreground object coated as white and the rest of the pixels are coated in black to easily detect the motion features in the image sequence. This becomes beneficial for the user as the colors are shown in contrast. The movements made with the hands are been tracked from the skeletal joints accordingly.

4. Contour extraction

Technique employed for Boundary following which is invoked in digital images in order to extract the boundary of the hand region and the variants made in fingers can be recognized with the perception of angles between the fingers. It is highly helpful in classifying the patterns through its accuracy and the computation time gets reduced over the consideration of whole pattern.

C. Enabling Gesture Recognition

After the detection of skeletal joint movements then need to train the system with the dataset that likely to be some verbal Gestures. Training being made by interpreting respective output (sign language words) for respective gestures. The trained gesture is been stored in the database in need to recognize them accordingly in which of how it has been trained. Then, the gestures are being recognized if the datasets are seemed to be matched.

D. Generate text for recognized gestures

Next task is to display the recognized gestures from the most matched interpreted dataset. The result will be enclosed through PC.

E. Synthesize Speech for the Text

The application was made finally after representing the recognized gestures as audible form. It can be carried out using utterance composed of words and phonemes that makes the sense after every accurate matching text.

IV. Results

The gestures shown in front of the camera are taken to the database to match across the dataset. If it matches over a certain range of dimensions, the respective text is generated and displayed. Further, the text is converted to speech using an API for converting text into speech. Following are the results of the processes.

V. Conclusion

A system for improved sign language recognition that involves procedures such as object detection, segmentation, background subtraction, motion detection, thresholding and contour extraction is proposed in this paper. The approach here uses a Kinect sensor to recognize gestures. Deaf and vocally impaired people can use the sign language recognition system to communicate with other people easily. The recognized gesture is matched with the corresponding text that generates and displays the assigned text. The strength of this approach is ease of implementation, no complex feature calculation and gives the result with higher recognition with less computation time..