

Hand-Gesture Recognition for Automated Speech Generation

Ujjayan Dhar¹ Sunny Patel² Suraj Gangwani³ Rohit Lad⁴ Pallavi Ahire⁵

^{1,2,3,4,5}Department of Information Technology

^{1,2,3,4,5}Sinhgad Institute of Technology, Lonavala, India

Abstract— In the past few years the development in the field of Gesture Recognition and the tools, devices, models and algorithms are into existence. Regardless of various innovations in the above technologies humans are still comfortable with showing gestures according to the plain hand. In this paper we focus on a system that runs on a mobile computing device which provides us the technology for automated translation of the Indian Sign Language system into Speech in the English Language and thus provides a duplex communication between the vocally impaired people and the common people of the society. This system can be very effective and can be used in the near future for communication between the people who do not understand the Sign language since it works on the principle of Gesture Recognition. The system uses an intrinsic mobile camera for gesture recognition and acquisition; gesture acquired is processed with the help of various algorithms. The system is able to recognize one handed sign representation of the standard alphabets (A-Z) and numeric values (0-9). The output of this system is very efficient, consistent and of high approximation of gesture processing and speech analysis.

Key words: Gesture Recognition, Human Computer Interactions (HCI), Sixth Sense Technology, Blob Detection, EmguCV

I. INTRODUCTION

In the recent advancements over the past years related to the Gesture Recognition, various technologies that were used in the latter years mostly depended on parameters like the use of multiple sensors or various hand wearable devices. Recently a new design epitome is being parturient in the domain of Gesture Recognition. As discrepant to the customary methods of capturing gestures with the help of various sensors like the 3- Axis Accelerometers in the KHU-1 Dataglove, usage of bend sensors[1], Hall effect sensors by assembling with the hand wearable device like the smart gloves, it is rather becoming very mainstream with the design paradigm[2] where gestures of the plain hand is directly captured with the help of an intrinsic camera that is available with the computing device and the respective event is triggered. Such Gesture Recognition paradigms like “Action-To-Speech”[3] where speech signals are recognized and the respective cue is being displayed and the even vision based sign language translator which is developed using the LABVIEW software shares the same concept as mentioned above [4].

Since gestures play a very important role thus the usage of gesture recognition can very be applied in various sections of the society, as its application ranges from the scientific usage, Human Computer Interactions (HCI) and even for the revivification of society. Sign Language is an abstract entity directed at natural language whose origin depends on the “Sign” or the “Gestures” and is the natural way for communication between the taciturn and vocally-debilitated people. Since various devices have arrived in the

market for the hearing impaired, they can listen the respective voices and thus they can communicate easily with each other. The paramount communication medium for the speech impaired people is the Sign Language. The pre-eminent issue arises when the people who are volitive with the speech impaired people or vice-versa. Since the one who is vocally impaired cannot respond to the conversations of the abled person, it creates a grave situation where the communication barrier materializes. To avoid such situation the person who is abled must have the knowledge of the Speech impaired standard gestures or must have a person who has the cognition of the language (translator). Thus the vocally impaired person can perform the Sign based language and the translator can translate into the Standard English language. But the translator cannot be carried at different locations as it is not every time practically possible.

The main paradigm that we focus on is to endeavor the linking between the Sign Language medium with the Standard English Language and thus providing the communication between the two communities in a seamless experience. The main concept simply focuses on the gesture capturing using mobile computing device. Initially the gesture which is to be captured is added in the Training Database and the respective word is being stored. When the gesture is being displayed in front of the inbuilt camera of the mobile device, it captures and attempts to map with the gesture that is stored in the Training Database, the mapping procedure of the gesture is done using various OpenCV algorithms. After the particular image has been mapped, by matching the contours of the stored image in the database, the respective word is being triggered and the finally the speech to voice conversion is performed.

II. LITERATURE SURVEY

Plain Hand Gesture Recognition system are an evolved form of the Sensor based hand wearable devices using sensors like the KHU-1 Dataglove [1]. Here hardware plays an important role in capturing the gesture and converting it into an 3-D model which processes the gesture and triggers the respective function. In the advancements of the wearable glove Bend sensors, Hall Effect Sensors and accelerometer is used and the data is them transmitted to the computer using the ARQ error preventing scheme [2]. Flex Sensors and accelerometers when assembled together provided the gesture recognition of the vocally impaired community [5]. The important tools and algorithm that played the important role in the Gesture Recognition is the Hidden Markov Model (HMM), even LABVIEW software [4] was the proposed system which can be incorporated in the mobile computing device, but the image capturing is done with the help of the micro camera fitted at the top of the head worn wearable and is captures the gestures and transits it to the mobile device. As far as the important concept in the field of Human Computer Interaction (HCI), Sixth Sense Technology plays an important role where Similar Hand

Gestures are captured with the help of a specific device made for that usage a.k.a the Sixth Sense Device which consists of a camera that is fabricated with the mini projector that is connected with the mobile computing device. The device simply captures gestures of the fingers as they are fitted with color markers of the color Red, Green, Blue and Yellow but the main drawback of this prototype is the user must carry the device and the fingers must be adorned with the colored marker caps[6].

III. SURVEY OF PROPOSED SYSTEM

By observing the above literature survey of the systems present in the innovations related to gesture recognition, few changes that are needed to be taken care of, they are as follows:

- 1) Detecting gesture with the help of Sensors and Datagloves should be replaced with plain hand-gestures.
- 2) Finding and developing more advanced algorithms that accurately give the gesture information.
- 3) Identify design issues both electronic and algorithmic and reduce the code complexity.
- 4) The latest systems should be developed according to machine learning concepts like Training and Testing Datasets.

The proposed system can be used to scan the plain hand gestures which follow the inbuilt Indian Sign Language System with sufficient amount of Training Datasets. The proposed system is helpful for the user to scan their gesture as it has a good amount of the gesture database (training datasets), hence improving the gesture to word conversion.

IV. SYSTEM ARCHITECTURE

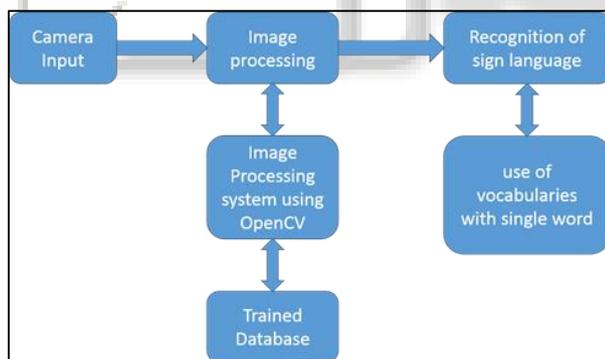


Fig. 1: System Architecture

Simple plain Hand Gestures are found to be efficient against all Wearable Devices. Thus they are likely to affect the way they work in a slightly different manner. The basic concept of our method is to capture gestures with the help of the intrinsic web camera. The system works on image processing algorithms using OpenCV Library and the Speech and Text Synthesizer. Initially the input hand gesture is given through camera in the RGB format in the Original Image block. Further the image is processed with the help of various image processing algorithms. OpenCV libraries are used in image processing system. These libraries are pre-defined in OpenCV. The trained datasets are used to compare with the input hand gesture image. Then finally if the given gestures are matched then the predefined word or sentence is visualized at the output textbox and with the help

of Text-To-Speech Synthesizer the following word or sentence is spelled. Finally the output is generated through text and speech format is generated as output.

V. MATHEMATICAL MODEL

A. Set Theory Analysis:

Let us consider two instances in the set dimension as the Video and Image points denoted by 'V' and 'I' and the Image Sample Set be denoted by consisting of the finite number of co-ordinates.

$$S = \{(V_1, I_1), (V_2, I_2), (V_3, I_3), \dots (V_n, I_n)\}$$

After V and I have been mapped the respective text values are triggered and denoted by "T" and converted to audio values denoted by "A"

Thus when we consider the sample sets generated for each instance points, it is denoted by

$$S = \{(V_1, I_1, T_1, A_1), (V_2, I_2, T_2, A_2), (V_3, I_3, T_3, A_3), \dots (V_n, I_n, T_n, A_n)\}$$

Identify the Functions as 'F'

$$F = \{F_1(), F_2(), F_3(), F_4()\}$$

F1 (V): Video Streaming.

F2 (I): Image Input.

F3 (T): Translated Text or Sound.

VI. CONCLUSION AND FUTURE WORK

Due to the advancements of the system initially starting from the usage of the Dataglove and sensors to the free hand gesture. The disadvantage that is mostly found in the system is due to the lack of the training datasets and the improvising of the speech synthesizer which is used for proper pronounce of the gesture-to-word input. The future scope of this system can be a system that can be designed for the blind people. In this system the camera can be used to capture the objects that are present in front of the user but not visible to the blind people. Thus the camera from the computing device can give the user the information about the distance of the street or the presence of any depressions on the road and even can be used to give the user the clean virtual view if the obstacles present in the path with the help of voice synthesizer.

ACKNOWLEDGEMENT

In this paper, we thank Prof. Pallavi Ahire and Prof. Rakesh Badodekar for their valuable information, feedback and comments. We would also like to thank Prof. N.A. Dhawas, Prof. S.D. Babar and Prof. T.J. Parvat who have helped us immensely and contributed their time and idea to this research

REFERENCES

- [1] Ji-Hwan Kim and Nguyen Duc Thang, "3-D Hand Motion Tracking and Gesture Recognition using a data glove", on Industrial Electronics ISIE 2009, July 2009.
- [2] Tushar Chouhan and Ankit Panse, "Smart Glove With Gesture Recognition Ability For The Hearing And Speech Impaired" on Global Humanitarian Technology Conference, September 2014.
- [3] Ibrahim Patel and Dr. Y. Srinivas Rao "RealTime Speech Processing For Automated Cue Generation", on ICCCT 2010.

- [4] Yellapu Madhuri and Anitha.G ,“Vison Based Sign Language Translation Device ”.
- [5] Neela Harish and S. Poonguzhali, “Design And Development of Hand Gesture Recognition System For Speech Impaired People”IEEE 2015.
- [6] S.Sadhana Rao,”Sixth Sense To Technology” on Communication and Computational Intelligence, December 2010.

