

Voice of Dumb

Dheeraj Kaushik

*Student, Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology, New Delhi, India*

Nilendu Ghosal

*Student, Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology, New Delhi, India*

Abstract— Communication is the man's most effective way to connect with the world around. But for those who lack the ability of speech, communication is visual, not verbal. Technology is being used at its best to make their lives easier and better in every aspect. Now days, Human Computer Interface (HCI) is evolving so as to simplify the interaction between man and machine; Gesture Identification being the latest technology to be explored. Modern gadgets like iPods, smartphones, laptops, and tabs etc. use gesture controlled technology to reduce the complexity of user interface. In this paper, an intelligent gesture recognition system is proposed which is less complex as it uses simple ways to detect hand gestures (without the use of color markers or sensors). This work aims to enhance the communication capabilities of the differently abled (speech impaired) people by analyzing the intricacies of Sign Language and implementing it in software.

Keywords—Human Computer Interaction (HCI); Gesture; Chironomy; Contour; OpenCV; Python; eSpeak.

I. INTRODUCTION

Image Processing is a signal processing technique that was introduced long ago in 1960s [1] and is still one of the red-hot topics for R&D in Industries. It deals with digitized images in order to extract relevant information and improve the quality of images. Different stages of image processing are shown in following figure:-

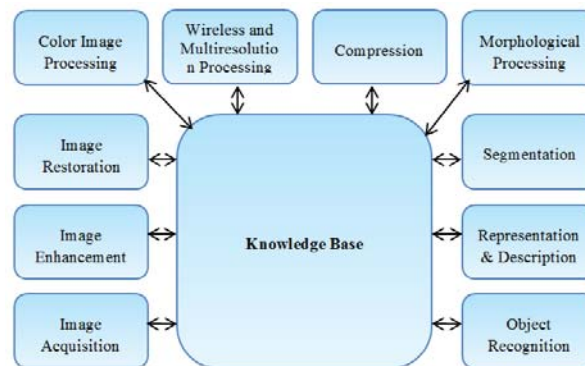


Figure 1. Stages of Image Processing [2]

Image processing finds its application in many areas like: - Office automation, Criminology, Biomedical, Astronomy, Robotics, and Gesture control etc. These days, gestures play a vital role in HCI. The most natural and absolute way to communication with differently abled people is through sign language (chironomy) [3]. Hence, we came with an idea of implementing the chironomy into hardware so that the gestures of differently abled people can be converted to voice output so that they can communicate efficiently with normal people thereby eliminating the need of understanding sign language. Several software such as python, MATLAB, C etc. are preferred for this purpose. To accomplish the task required for this project, we are using python along with openCV.

In the coming sections, we will describe the implementation aspects of our project. The details of the software used is given in the below section.

II. SOFTWARE DESCRIPTION

The proposed project, “Voice of Dumb” uses following software for real time image processing: openCV, numpy, eSpeak and python. Each of the software used is explained below:

- **OpenCV:** OpenCV was originally introduced by INTEL. It is an open source library that is used for image and video analysis [4]. It is written in C and C++ and provides support for LINUX, Windows and MAC OS. The major changes in this library took place in 2009 and named as OpenCV2. We have used OpenCV2 to carry out our task. Some of the platforms that widely uses openCV for image processing are C, C++, python, MATLAB, ruby etc. There are more than 3000 high level computer vision (CV) algorithms are available that significantly helps programmers to develop optimized programs.
- **Numpy:** This library is written in C and helps in scientific computations. It consists of N-dimensional array objects that are used in our project for array arithmetic operations. The elements of this N-dimensional array must be of the same type. Numpy maintains the high abstraction level of Python as well as efficiently exploits the performance of C using the array operations [5].
- **eSpeak:** eSpeak is the library that converts text input to corresponding voice output. This library is integrated with other libraries to generate voice output of corresponding hand gestures. It consists of voices which can be modified in loudness, pitch, tone etc. as per the user’s need [6].
- **Python:** During the past fifteen years, Python has emerged as the fundamental standard for interactive and computation-driven scientific research [7]. It supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. It supports a wide range of libraries and helps the user to reduce the lines of codes. All the libraries discussed above are integrated by using python.

III. DESIGN AND IMPLEMENTATION

The below diagram explains the steps involved in the project.

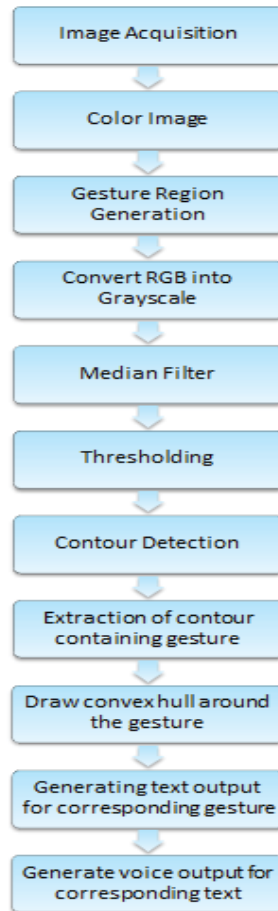


Figure 2. Flow Chart of our project

- **Image Acquisition:** It is an image processing action of retrieving images from source for processing.
- **Color Image:** The image taken by the camera will be a color image with various unwanted information.
- **Gesture region Generation:** To reduce the background noise, we have formed a gesture region in which user can make gestures.
- **RGB2Gray:** The color image acquired is converted into grayscale for further image processing.
- **Median Filter:** For smoothening the acquired image, blurring is performed by using median filter.
- **Thresholding:** This operation is performed to detect the frontal images and eliminate the background images by selecting the proper threshold value. Threshold value can vary with the amount of brightness in surrounding.
- **Contour Detection:** Contour detection is the method of detecting closed figures present in our image.
- **Extracting Contour containing hand gesture:** To extract contour containing hand gestures, we will calculate the area of all the contours detected in the image and since we have made a gesture region so the contour with maximum area will be our hand.
- **Draw convex hull:** In this step, we will draw the convex hull of our gesture extracted in the previous step.
- **Generating the text output:** On the basis of number of defect points, text output is generated for the corresponding gesture.
- **Generate voice output:** With the help of eSpeak library, the generated text is now converted to a voice output.

Above points are used to carry out our project with all the positive results.
The gesture recognition results are shown below:

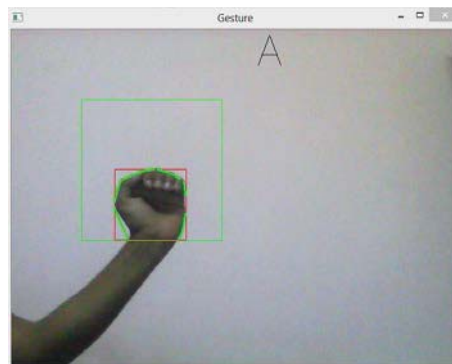


Figure 3. Gesture for alphabet, A

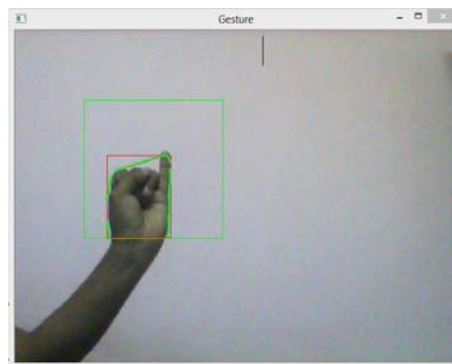


Figure 4. Gesture for alphabet, I

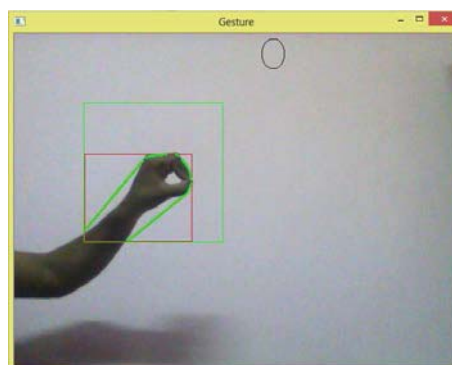


Figure 5. Gesture for alphabet, O

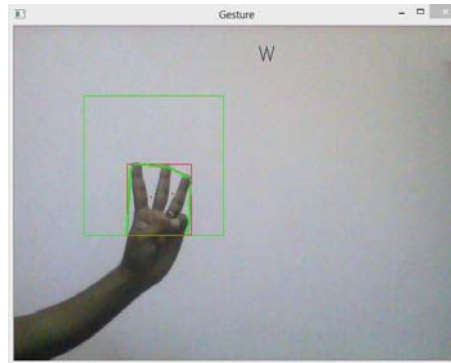


Figure 6. Gesture for alphabet, W



Figure 7. Gesture for the word, VICTORY

IV. CONCLUSION

In this paper, we have successfully come up with an intelligent solution for communication problem of speech impaired people. The method proposed for interaction is found to be effective and efficient with least complexities and yields satisfactory results.

During the testing phase of our software, we encountered problems such as background noise and delay of voice; also we found that the performance of the software is directly depends on the quality of camera. To overcome these problems and achieve optimum performance, we have used background noise reduction techniques (discussed in Section III) and high resolution camera.

Furthermore, these bugs will become negligible and the performance will increase when we take this software to the next level i.e. dedicated hardware implementation.

V. FUTURE SCOPE

In the nearer future, we will take this idea one step forward and implement it on hardware so that it becomes a practically usable product. The product will use high speed microcontroller (for processing), a camera (for hand gesture recognition) and a speaker (for generating speech output). In all, the proposed product will be compact, cost effective and easy to use.

REFERENCES

- [1] Ashraf A. Aly, Safaai Bin Deris, Nazar Jaki, "Research review for Digital Image Segmentation techniques," International Journal of Computer Science & Information Technology (IJCSIT) vol. 3, No 5, Oct 2011.
- [2] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing," 3rd ed., page 26.

- [3] Luo R.C., Yen-Chang Wu, "Hand gesture recognition for Human-Robot interaction for service robot," MFI, 2012 IEEE Conference, Hamburg, pp. 318-323, Sept. 2012.
- [4] Culjak I, Abram D, Pribanic T, Dzapo H, Cifrek M, "A brief introduction to OpenCV," MIPRO, 2012 Proceedings of the 35th International Convention, pp. 1725-1730, May 2012.
- [5] Mads Ruben Burgdorff Kristensen, Brian Vinter, "Numerical Python for scalable architectures," PGAS'10, ACM Proceedings, New York, 2010.
- [6] Mentor Hamiti, Ramiz Kastrati, "Adapting eSpeak for converting text into speech in Albanian," IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, July 2014.
- [7] K. jarrod Millman, Michael Aivazis, "Python for Scientists and Engineers," IEEE Computing in Science & Engineering, vol. 13, issue 2, pp. 9-12, March-April 2011.