

Performance and Detection of Gesture and Motion for Deaf and Dumb Communication

Shoaib Ibrahim Shaikh¹, Ismail Mehmood Memon², Sanmay Jayaram Shetty³,

Aziz Saifuddin Vakanerwala⁴, Supriya H. Madane⁵, Aniket Suresh Ingavale⁶

Student, Dept. of Information Technology, Trinity College of Engineering and Research, Pune, Maharashtra, India¹

Student, Dept. of Information Technology, Trinity College of Engineering and Research, Pune, Maharashtra, India²

Student, Dept. of Information Technology, Trinity College of Engineering and Research, Pune, Maharashtra, India³

Student, Department of Information Technology, Trinity College of Engineering and Research, Maharashtra, India⁴

Assistant Professor, Department of Information Technology, Trinity college of Engineering and Research, Pune
Maharashtra, India⁵

Assistant Professor, Department of Information Technology, Trinity college of Engineering and Research, Pune
Maharashtra, India⁶

ABSTRACT: Communication is the most important aspect in today's world. Being social helps us to grow and understand people. But communication is only possible when there are two or more people who have the necessary means for communication. Today's technology has made communication possible over very long distances, but the basic need for communication is the power of hearing and speech. This is unfortunate in the case of some people (i.e. deaf and dumb). As we know that the deaf and dumb communicate using sign languages, we can use technology to make the most of this ability for communication. In this paper, we have proposed a dynamic gesture recognition system which will identify the letters of the English Language and convert them to text and speech. The major steps involved are Video Acquisition, Pre-processing, Color Conversion, Color Localization, Template Processing and Gesture Detection. This system will help in making communication between deaf and dumb people and normal people possible.

KEYWORDS: Communication, social, dynamic, gesture, motion, recognition.

I. INTRODUCTION

In the last several years a lot of R & D has been done in the field of gesture and sign language recognition [1]. This has given rise to Human Computer Interaction and Intelligent Computing. HCI focuses primarily on the interaction between the people and computers. Researchers and programmers study the ways in which a normal user interacts with the computer system. This observation motivates the programmer to develop systems in such a way that make this interaction as simple and as convenient as possible for the users. We study HCI to understand how we can make the life of a computer user less complicated using computer technology. In many applications, gestures are an important aspect around which a programmer develops a system that involves Human Computer Interaction. Most systems involve gestures as input to give a particular output. Systems like sign language communication [1], virtual reality and robotics greatly rely on gestures as input. Gestures consist of two parts namely static gesture [2] and dynamic gesture [3]. Static gestures are described on bases of hand shapes and dynamic gestures using hand movement. For implementation of gesture recognition system requires camera, data gloves [4], Color markers for capturing and tracking gestures.

The objective of this paper is to develop a gestures recognition system which will help deaf and dumb communicate using gestures. According to the American Sign Language (ASL), each gesture of the ASL is a representation for a word in the English Language. So we know that the deaf and dumb are aware of the English language and they can

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

easily understand it. This system will identify the words of the English language when the dynamic gesture of a letter is made in front of the webcam. For hearing impaired, sign language is way of communication which employs different sign, gesture made by moving the hand. So we are developing such a system which can recognize this hand movement. Our work will be in the field of dynamic gestures. Our project uses a live video feed to capture the pattern, gesture that is generated in front of the webcam [1]. A Color marker will be used to identify the gesture.

II. RELATED WORK

A. Vision Based Approach

Vision based approach [2][3][4][5], system requires only camera to capture the image for human interaction. In this approach processing is done on already stored image. Arpita Ray Sarkar and G Sanyal developed a hand gesture recognition system which consist of models like data acquisition, gesture modelling, feature extraction and hand recognition. In this paper image is captured using input device which is used in data acquisition. Then gesture modelling is used which consist of hand segmentation, thresholding, Color normalization, salt and pepper, morphology erosion, multidimensional mean and edge detection [8]. And the features are extracted in order to recognize the hand. In this paper Feature extraction uses hidden Markov model. These system are simple but faces a lot of challenges like complex background, lighting variation and skin color.

B. Data glove approach

Data glove or hand glove use sensor devices for capturing hand motion [5][6][7][9]. Glove provides exact co-ordinates of palm and finger's location, orientation and hand configuration. V. Padmanabhan, M. Sornalatha developed data glove based hand gesture recognition system, which has models like flex sensor, hand gesture, accelerometer, gesture recognition, display unit, speech synthesis [19]. Five flex sensors were used on each finger to measure the x, y and z position. This coordinate are used in recognition the gesture and its related letter. This data is sent to speech synthesizer to form the speaker, speak that individual word.

In another paper by Piyush Kumar, Jyoti Verma and Shitala Prasad uses VHand 2.0(5DGT), AD convertor, Bluetooth protocol, Mapping Function for human computer interaction [6][7][10][11-13]. This approach was used for Google Earth manipulation, and sketching purpose in OpenGL 2D Graphics Environment. However, data glove approach requires the user to be connected with the computer physically. This increases the user computer interaction, hence obstacle the ease of use. Beside purchase cost is required.

III. SYSTEM ARCHITECTURE

In our approach, user will be using a color pointer, so that we get an accurate pattern. This paper consists of the detailed concepts of image processing, human computer interface, conversion of gesture to text and text-to-speech [11]. This system is going to be beneficial for the people who are devoid of hearing and speech.

The proposed system consists of following units, namely:

1. Video Acquisition
2. Pre-processing
3. Color Conversion
4. Color Localization
5. Template Processing
6. Gesture Detection

1. Video Acquisition:

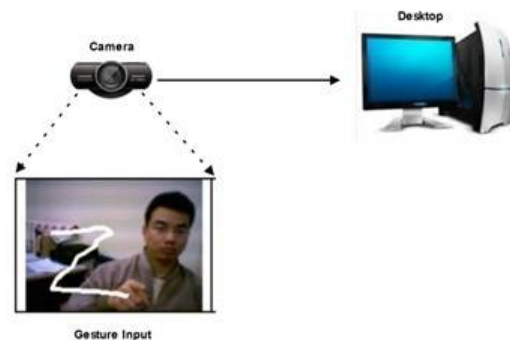
1.1 Video Input and Frame Extraction

We develop a system which will capture gesture by using in-built web cam of a laptop or an external web-cam connected to a PC. The resolution of this webcam has to be of 320x240 pixels so that gesture will be captured properly.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016



(a)

Figure (a).Shows the Block diagram of gesture and motion recognition system

Webcam is used to capture image or as communication medium over the internet. Webcam is intergraded with computer system with the help of driver and libraries which help to make use of webcam. These drivers are built according to processor and Operating System. Our first step will be to integrate this webcam function into our system to capture gestures. For this purpose we will be using JMyron library. We will capture live video feed and display on screen.

2. Pre-processing:

2.1 Gaussian blur

Gaussian blur is used in image processing to reduce noise. The graphical effect of this blurring technique is a smooth image which has a similar appearance to that of viewing the image through a translucent screen. Images are not smooth because neighbouring pixels are distinct. Blurring makes neighbouring pixels look more similar. Using this smoothened image we can distinctly identify relevant pixels (i.e. Gesture pixels) from irrelevant pixels.

Formula: $\text{Sum R} = (R_1 + R_2 + \dots + R_9) / 9$

$\text{Sum G} = (G_1 + G_2 + \dots + G_9) / 9$

$\text{Sum B} = (B_1 + B_2 + \dots + B_9) / 9$

3. Color Conversion:

3.1 RGB to HSV

The web-cams in all devices use the RGB color model. We will be converting the input from RGB color model to HSV color model. The reason for this conversion is because the HSV model is a stronger model than the RGB color model as it offers a more intuitive representation of the relationship between colors. HSV is described using color[10], vibrancy and brightness. HSV model depicts how our eyes view colors. These values are then used to determine particular area of region.

$\text{RGBmax} = \max(R, G, B)$, $\text{RGBmin} = \min(R, G, B)$

$V = \max(R, G, B)$

$S = 255 * ((\text{RGBmax} - \text{RGBmin}) / \text{RGBmax})$

a) If $\text{RGBmax} = R$, then

$H = 0 + 43 * ((G - B) / (\text{RGBmax} - \text{RGBmin}))$

b) If $\text{RGBmax} = G$, then

$H = 85 + 43 * ((B - R) / (\text{RGBmax} - \text{RGBmin}))$

c) If $\text{RGBmax} = B$, then

$H = 171 + 43 * ((R - G) / (\text{RGBmax} - \text{RGBmin}))$

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

3.2 Thresholding

The normal HSV image is converted into a binary image by thresholding. This step is important as the image is converted to black and white so as to obtain a distinct image of the pattern. The white part in the obtained image will be the pattern whereas the black part will be the background which will be neglected. This step will generate the simplest form of the image and will be easy to map it to our databases which contain the pattern and their outputs.

4. Color Localization

4.1 Blob detection and filter

A Blob is a cluster of connected pixels in a picture that share some common property (Eg: grayscale value). The goal of blob detection is to spot and mark the dark region in the image. The system starts scanning from the top and searches for the first black pixel. As soon as it finds the first black pixel, it will save its 'x' and 'y' co-ordinates. Using these co-ordinates, the maximum and minimum values of x and y co-ordinates will be saved into vectorentry table. This process will be repeated till the entire image is scanned and all the related blobs are found. These points will be joined in order to form a blob [20].

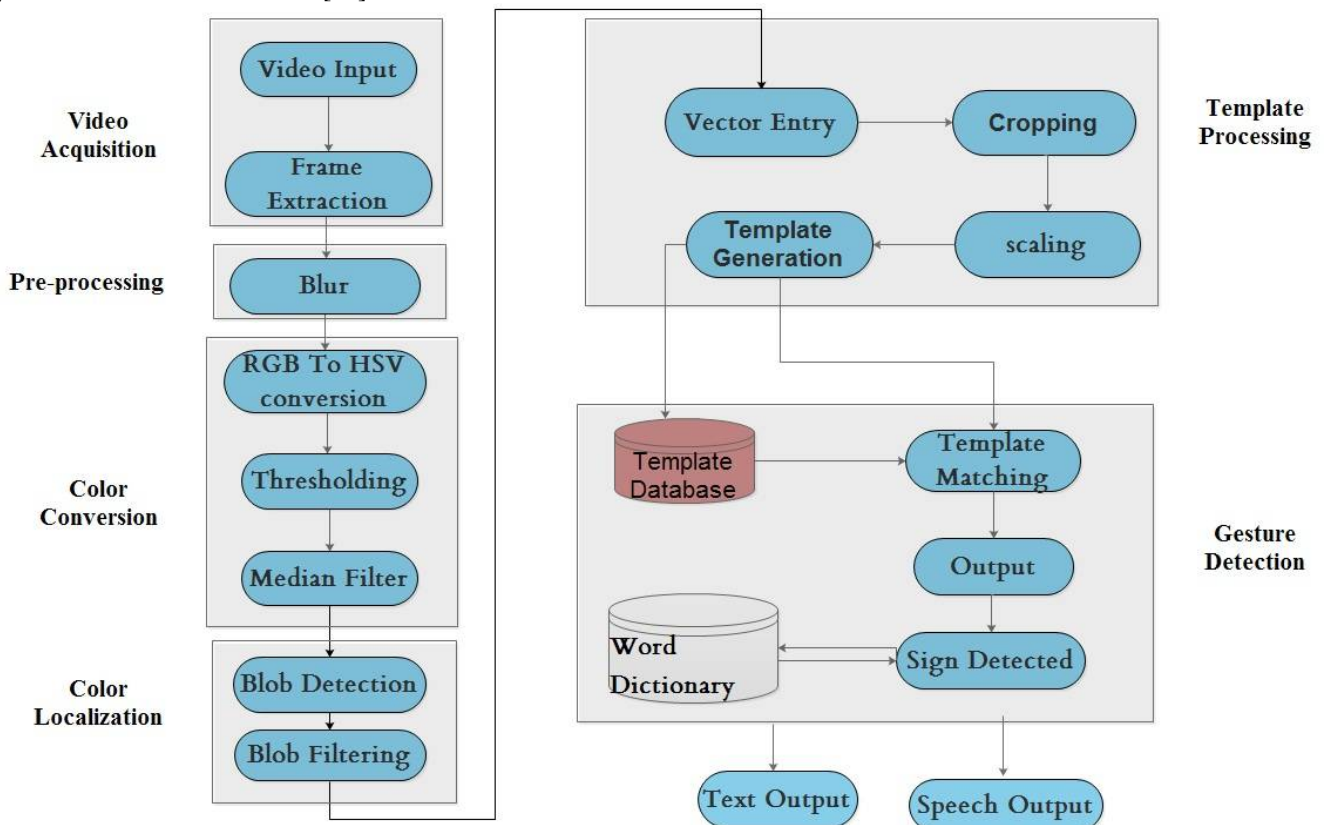


Figure (b). System flow diagram of gesture and motion recognition system

The major steps involved are Video Acquisition, Pre-processing, Color Conversion, Color Localization, Template Processing and Gesture Detection.

5. Template Processing

5.1 Cropping and Scaling

Cropping is taking out the relevant part of the image and discarding the rest of it. Cropping makes use of the maximum and minimum values of x and y.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

Scaling is used to resize the cropped frame to a required size. The input image is resized to certain size of pixels(eg: 100x100 pixels). This is done to reduce the processing time by laptop and to improve the accuracy of the system.

5.2 Template Matching using Euclidean Distance

By using above techniques, multiple templates are generated and each gesture has its own template. These templates are then compared with our database system to find a matching template.

We use the Euclidean distance [17], for calculating the minimum distance between the input image and the image to be recognized from the database. If the distance is small, we say the images are similar and we can decide the most similar image in the database. This template is then converted into text. This step is called template matching.

6. Gesture Detection

6.1 Text –to-Speech API

There are many existing API, which can be used in our system. These API's are capable of converting Text to speech. We are going to make use of any one of such API in our system which will read the text that will be formed after the gesture or pattern is correctly recognized. We will be using Windows SAPI as Windows is widely used platform.

6.2 Word Dictionary

User makes gesture for each and every letter. Due to this, it will become hectic for the user to make gesture for the entire sentence. To reduce this, we will maintain a word dictionary which will help the user in sentence formation. When sign is made then that letter is converted to text. Then that text is searched in the dictionary letter by letter, the words in dictionary are stored alphabetically. The words matching the text will be suggested. If the suggested word is what the user is looking for then a gesture is made by the user to complete the word, this reduces the time and effort of the user to make gestures of the remaining letters of the word.

IV. EXPECTED RESULTS AND FUTURE SCOPE

In our proposed system, the pre-processing modules are capable of converting any image into a black and white image where the black part of the image will determine the unwanted section of the image whereas the white part will be the important section of the image. All these images will be merged together to form the letter that will be made in front of the webcam. Once this is recognized, it will be matched with our database and identified. The expected result of our proposed system will be 85-90%.

Our proposed system has been made to support English language. This system can be created for other languages as well. Also, there are many applications that can be modified to be controlled by gestures. Such application can also be created using the technology of our system. Robots and drones can also be configured to be handled with the use of gestures. This system will make the handling of such gadgets easy.

V. CONCLUSION

In this paper, a dynamic gesture will be made in front of a webcam that will be a letter of the English language. This gesture will be made by a color pointer placed on the finger and the letter will be drawn in front of the webcam in air. Gaussian Blur algorithm is performed on the frames to make the image smoother. RGB to HSV conversion will give us a distinctive image. HSV thresholding will give a black and white output of this image. Blob Detection is used to detect the position of the region where the concerned part of the image is situated. All these frames will be merged again and the image of the letter will be in white while the entire background will be black. Template Matching will be used to match this image with our database. These steps will be repeated until the user finishes giving input and the final result will be displayed on the screen. Additionally, there is a speech output that will read out the word or sentence that has been displayed. Moreover, to make this even more convenient, we will add a data dictionary that will suggest words so that time taken for making gesture for the entire word will be saved.

Using this system, we are trying to help achieve easy communication between the normal people and the dumb and deaf to bridge the gap between them so that they live their lives normally and express themselves in a much better way.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

ACKNOWLEDGMENT

Sincere gratitude to our project guide Prof. Aniket Suresh Ingavale, Supriya H Madane(I.T Department) Trinity College of Engineering and Research, Pune, for their valuable time, support and guidance during the course of the paper. We take this opportunity to convey our sincere thanks to all the individuals who have assisted and helped us in carrying and bringing out this paper. Last but not the least, the support and help of all the teachers and staff members of Dept. of I.T, is gratefully acknowledged.

REFERENCES

- [1] Beifang Yi, Xusheng Wang, Frederick C. Harris, Jr, and Sergiu M. Dascalu "sEditor: A Prototype for a Sign Language Interfacing System" Human-Machine Systems, IEEE Transactions on (Volume:44 , Issue: 4), pp 499 – 510,2014.
- [2] Chen YongQiang, HUANG YingShu, Xia Chenjie "Static hand gesture recognition algorithm based on finger angle characteristics", Proceedings of the 33rd Chinese Control Conference, pp 8239 - 8242, 2014.
- [3] JoyeetaSingha, and Karen Das, "Recognition of Indian Sign Language in Live video", International Journal of Computer Applications", Vol.70, No.19, pp 17-22, May 2013.
- [4] PragatiGarg, Naveen Aggarwal and SanjeevSofat "Vision Based Hand Gesture Recognition", World Academy of Science, Engineering and Technology 49, pp. 972-977, 2009.
- [5] G. R. S. Murthy & R. S. Jadon, "A Review of Vision Based Hand Gestures Recognition", International Journal of www.ijcat.com 17 Information Technology and Knowledge Management, vol. 2(2), pp. 405-410, 2009.
- [6] Shitala Prasad, Piyush Kumar, KumariPriyanka, "A Wireless Dynamic Gesture User Interface for HCI Using Hand Data Glove" ,Contemporary Computing (IC3), 2014 Seventh International Conference, , pp. 1617-1621,2014.
- [7] Piyush Kumar, JyotiVerma"Hand Data Glove: A Wearable Real-Time Device for Human-Computer Interaction" International Journal of Advanced Science and Technology Vol. 43,pp-15-26, June, 2012.
- [8] Arpita Ray Sarkar and G sanyal "Hand Gesture Recognition Systems" International Journal of Computer Applications2013 by IJCA Journal Vol. 71, No. 15,pp 25-37,2013.
- [9] V.Padmanabhan, M.Sornalatha, "Hand gesture recognition and voice conversion system for dumb people" International Journal of Scientific & Engineering Research, Volume 5, Issue 5,pp 427-431,May-2014.
- [10] YogeshTayal, RuchikaLamba, SubhransuPadhee, "Automatic Face Detection Using Color Based Segmentation", International Journal of Scientific and Research Publications, Volume 2, Issue 6,pp 1-7, June 2012.
- [11] Joseph J. LaViola Jr. "A Survey of Hand Posture and Gesture Recognition Techniques and Technology", Master Thesis, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA, 1999.
- [12] AnupamAgrawal, Rohit Raj ShubhamPorwal, "Vision-based Multimodal Human-Computer Interaction using Hand and Head Gestures", Information Communication Technologies (ICT), IEEE Conference, 2003, pp.1212-1217, 2013.
- [13] JinhuaZeng, Yaoru Sun, and Fang Wang, "A natural hand gesture system for intelligent human-computer interaction and medical assistance",pp 382 - 385 ,2012.
- [14] AditiKalsh,N.S. Garewal" Sign Language Recognition for Deaf & Dumb" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013.
- [15] "Euclidean distance matrices: Properties, algorithms and applications" Reza Parhizkar - 2013/2/7 Institution ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
- [16] Clifton Phua1, Vincent Lee, Kate Smith, Ross Gayler, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor", Australian Research Council, Baycorp Advantage, and Monash University LP0454077, 2008.
- [17] Pankaj Kumar, Anthony Dick and Tan Soo Sheng, "Real Time Target Tracking with Pan Tilt Zoom Camera", Digital Image Computing: Techniques and Applications", pp.492-497, 2009.
- [18] Robert Y. Wang, "Real-time hand tracking as user input device", Proceedings of 5th international conference on Knowledge, information, and creativity support systems, 2010.
- [19] S. Malik," Real-time Hand Tracking and Finger Tracking for Interaction", Tenth IEEE International Conference, 2012.
- [20] Robert T. Collins,Carnegie , "Mean-shift Blob Tracking through Scale Space" Mellon University,2003