

Alphabet Recognition for Deaf-Blind People

Yuvraj Singh Rathore¹, Charvi Mittal², Avik Basu³

¹(Electrical and Electronics Engineering, Vellore Institute of Technology, India)

²(Electronics and Communication Engineering, Vellore Institute of Technology, India)

³(Computer Science Engineering, Vellore Institute of Technology, India)

Abstract : This paper aims to present a system to aid deaf-blind people in communicating with the outside world efficiently and help ease their burden. They often find it difficult to interact with others and experience inability to perform daily life activities. The algorithm proposed in this paper recognizes various hand gestures performed by user in the form of alphabets in front of the integrated camera. These are tracked using image and video processing techniques and the character is further detected using pattern recognition. A feedback mechanism has been also employed for achieving high accuracy. The above algorithm was implemented using OpenCV 2.4.9 and Beaglebone Black. The proposed method is user friendly and it will translate hand gestures into meaningful characters with high efficiency.

Keywords: Object Detection, Object Tracking, Pattern Recognition, Real Time Image Processing

I. Introduction

Many real world applications require the use of techniques such as image processing, object recognition and motion detection. Extensive research has been done in image processing, computer vision and pattern recognition domain in the past few decades. These algorithms when implemented have several applications such as face recognition [1], obstacle detection [2], medical diagnosis [3], fingerprint recognition [4], body sensing and many more. Several algorithms have been devised and implemented upon for the automation of machines and providing them enough intelligence to work independently.

Deaf-blindness is a condition in which the combination of hearing and visual losses in children causes “such severe communication and other developmental and educational needs that they cannot be accommodated in special education programs solely for children with deafness or children with blindness” [5]. The challenge of learning to communicate is perhaps the greatest one that children who are deaf-blind face. According to Sense, a national charity in UK, there are more than 3,56,000 deaf-blind people in UK alone. Hence, this paper explores one of the ways to detect characters drawn using meaningful hand gestures which eliminates the use of keyboards or any other typing devices [6] and can be easily used by deaf-blind people for communicating with the outside world. It also forms an integral part of major future technologies and can be further improved and implemented in several more areas to increase efficiency of various devices and make them hands-free.

Pattern Recognition involves useful classification of input data using various feature extraction algorithms on noisy random input data [7]. For example, a pattern could be a fingerprint image, a handwritten cursive word and so on. Humans have highly developed skills for sensing and recognizing different patterns such as handwriting recognition, face recognition and understanding spoken words. Using Pattern Recognition, similar capabilities can be assigned to machines.

In this paper, an algorithm has been proposed in which hand gesture with the help of a colored object is performed in front of an integrated camera. The performed hand gestures are in the form of desired alphabet. In this algorithm, centroid of the tracked object is determined in each frame. Location of these centroids in all the frames is plotted and the locus is drawn. On this character, pattern recognition algorithm is applied to extract features. These features are compared with the features present in the database. The best matched feature among the database is returned and displayed on the screen. A feedback mechanism is deployed to maintain a high accuracy of the system. Hence, application of above illustrated techniques will help deaf-blind in communicating with the outside world successfully and efficiently.

II. Methodology

2.1 Input from camera

A camera is integrated with OpenCV 2.4.9 in order to capture real time frames. The integrated camera used in our system acquires frame at the rate of 11 frames per second. System can be operated in 3 different modes according to the choice of the user. During the initialization of the system, user will be asked for the mode of operation namely slow, medium and fast. According to the choice given by the user, integrated camera will be triggered and it will capture 25 frames if mode selected is fast, 50 frames for medium and 75 for slow. The default resolution of the input RGB image is taken as 1280x720 pixels.

2.2 Color Space

Frames captured from integrated camera are in RGB format. Now, in order to detect blue color, RGB frames are converted into corresponding HSV (Hue- Saturation- Value) color plane. HSV plane is preferred as it is less sensitive to lighting conditions and it uses only a single number to detect the color (hue) despite the presence of several shades of a particular color ranging from lighter to darker shades. [8] The purity or amount of color and brightness of the color are handled by the “saturation” and “value” parameters respectively.

2.3 Thresholding

The pixel values corresponding only to the blue object are extracted and consequently a range is selected in which different shades of blue exist. This range was found out after grabbing some screenshots of the colored object in which we were interested in tracking and observing the kinds of Hue, Saturation and Luminosity values that occur and hence found suitable maximum and minimum values. Following the above approach the range for HSV values that are bluish in color was found to be between (104, 178, 70) and (130, 240, 124).

2.4 Median filter

In order to smoothen the resultant image found at the step 2.3, median filter is applied over 3x3 pixels.

2.5 Removal of noise

Foreground in the above image will be represented by 1 and background as 0. All the connected components in the resultant binary image are labeled. Areas of all the labeled components are calculated and a threshold value of 2000 is selected. So, all the labeled objects whose area is greater than 2000 will be represented with white pixels or 1 while remaining objects will be represented with black pixels or 0. Noise generally occurs in smaller areas. Hence, effective noise removal is possible and only the region of interest is extracted[9].

2.6 Centroid calculation

For centroid calculation, firstly, it is required to determine the sum of X-coordinates and sum of Y-coordinates of all the white pixels and then count number of white pixels (N) present in the image. So, coordinates of the centroid can be given by $(\sum x/N, \sum y/N)$. In this way centroid of all the labeled components in each frame is determined and plotted. All the centroids are joined along a straight line to display the locus being formed on the graph plot.

2.7 Extraction of Region of interest

Depending on the distance of hand gestures from the camera the area covered by the foreground (white pixels) will vary. So, in order to avoid this discrepancy, region of interest which includes only white pixels is extracted and resized to 900x1200 pixels. Region of interest is extracted by determining the leftmost and rightmost white pixel's column and topmost and bottom-most white pixel's row. This image is then cropped about these particular rows and columns. This cropped image is then resized to 900x1200 pixels without changing the aspect ratio of the original image.

2.8 Pattern Recognition

Above plotted character will resemble an unknown pattern to the machine. So in order to understand this unknown pattern with significant accuracy, pattern recognition algorithm is employed on it. First of all, a database of patterns is created for every alphabet. Five arrays of each alphabet are produced to make the recognition process more robust. Above image is divided into 12 blocks or 4x3 grids. Now, in each block number of white pixels are computed and stored in an array. This array formed is compared with all such arrays for different alphabets present in the database. This comparison is achieved by calculating the root mean squared error (RMSE) of the obtained array with corresponding array for a particular alphabet in the database. Minimum error will correspond to the best match alphabet. The root mean squared error is determined by following formula:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x'_i - x_i)^2}{n}}$$

(1) [10]

2.9 Feedback Mechanism

In order to confirm that the gesture performed by the user is conveyed correctly, a feedback mechanism is devised. This feedback consists of DC motors which are powered by Beaglebone Black. The device is placed on the back of the palm. The movement of motor in the form of detected alphabet constitutes this feedback. As soon as the letter is detected, the feedback system is triggered, and motor movement is controlled accordingly.

2.10 2-Way Communication

In order to achieve effective two-way communication a keypad, which includes all the 26 alphabets, is interfaced with Beaglebone Black. The person who wants to communicate with the user will input the message alphabet by alphabet with the help of keyboard and it will trigger the feedback system discussed in step 9 to follow the pre-defined path corresponding to a particular alphabet. This pre-defined path followed by motor mimics the person's input alphabet on the back of user's palm. In this way user will be able to comprehend the other person's message with the help of its sense of touch.

III. Simulation Results

All simulations were carried out using OpenCV 2.4.9. Following is an example of user who is trying to convey the character 'Z' to the outside world. So user will perform hand gestures in front of an integrated camera in the form of 'Z'. Fig. 1 given below shows the RGB image of the user performing hand gestures with the help of blue object.

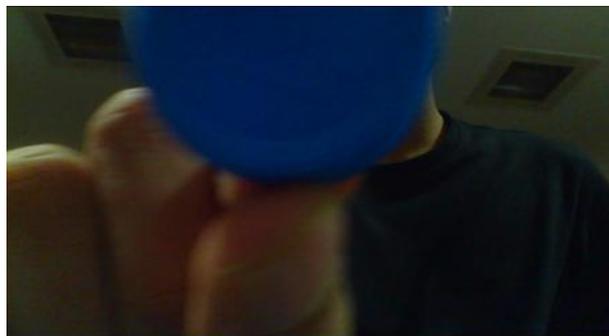


Fig. 1

Above RGB image is converted into the corresponding HSV image as shown below in Fig. 2.

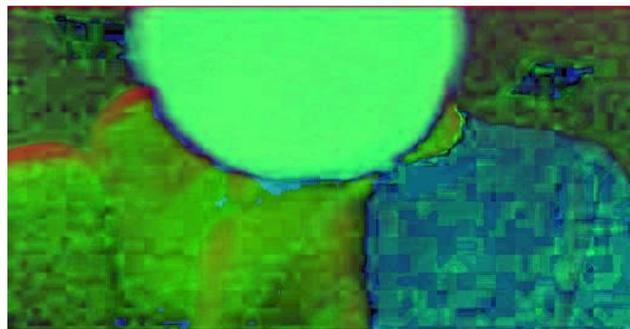


Fig. 2

After applying threshold on the above HSV image, it results in the following image shown in Fig. 3.

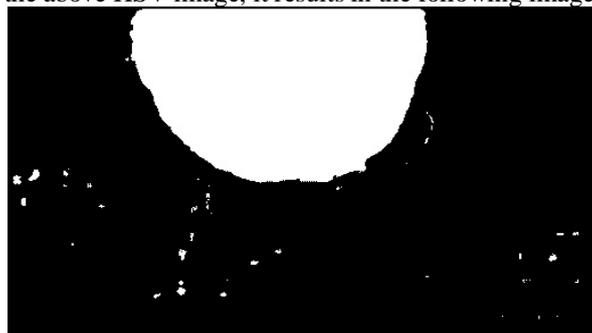


Fig. 3

In the above figure all small holes are removed after applying threshold [9] on the area covered by the white pixels and since noise is present in smaller areas small holes are filled as shown in the Fig.4 below:



Fig. 4

The centroid of the blue object which is represented by white pixels in Fig. 4 is detected in all the frames and plotted as shown in the Fig. 5.

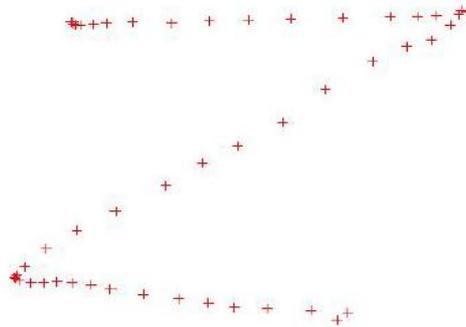


Fig. 5

These centroids are then joined by a straight line and resultant figure is shown below in Fig. 6.

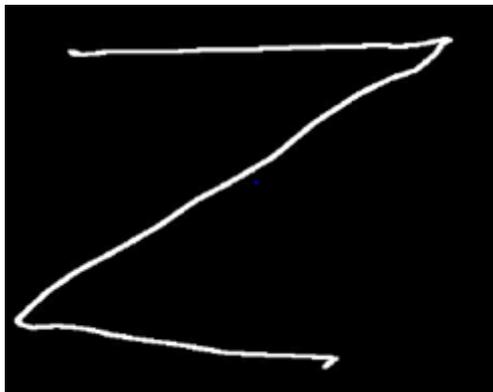


Fig. 6

Region of interest is extracted from the above image as explained in the step 2.6 of the methodology and then resized to 900x1200 pixels. Following resultant figure is obtained:-

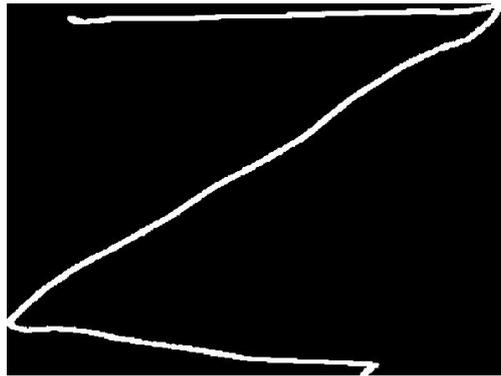


Fig. 7

This final character is now divided into 3*4 grids as shown in figure 8 given below and number of white pixels in each block is calculated. These white pixels calculated in each block are stored in an array. Now this array is matched with predefined arrays in the database and finally best match is determined.

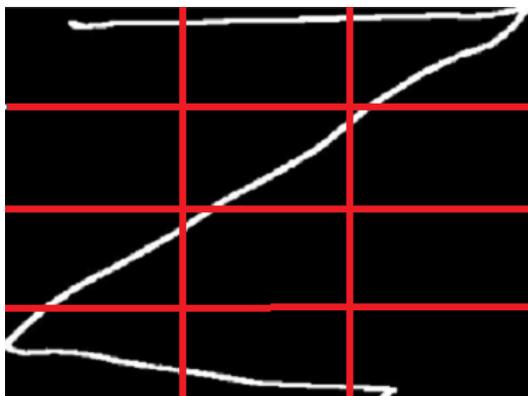


Fig. 8

1011100001111111	1010010101010101	1000001111111101
1001000101001010	0100101010101010	1000010100100101
1010101001010101	1010010101010010	0101001010101010
1111100010101000	1010100101010101	1100001010101001
1010101010101010	0101010101001010	0101010101010101
0101010101010101	1010010101011000	0010101010101010
1010100101010101	1100001010100111	1100100101011100
0101001010101010	1101001010100001	1010010101010101
0110010101000011	110111110010101	0001010111110010
1110001010101001	1110010101010010	1011010101010100
1101010111110011	1010101001011111	1011001010101011
1111010010101001	0010100101111110	1110001001010100

Fig. 9

The best matched alphabet is then sensed by the user with the help of feedback mechanism and if the user's response matches with the feedback's response, then the corresponding alphabet is displayed on the screen else the user will again repeat the whole process in order to convey the meaningful information.

IV. Conclusion

The proposed system is capable of translating meaningful hand gestures into characters with high efficiency. This assistive technology for deaf blind people is unique as it employs a rather simple and robust algorithm which is user friendly. This algorithm when tested with all the alphabets successfully detected correct gestures 97% of the times. In order to compensate for that 3% error, system also consists of a feedback system which will not allow the alphabet to be displayed on the screen unless user response matches feedback's output. Hence, it completely eliminates the probability of incorrect response being conveyed to the concerned person involved in two way communication with the user. Above mentioned algorithm's accuracy may be improved by increasing the number of blocks in which the image is sub-divided as explained in methodology in section 2.8. But this will also increase the computational time and system's execution time will be affected. According to the priority of the user for accuracy or execution time, number of blocks must be selected in which the image is to be divided. The system developed here will not only be beneficial to deaf blind but it can provide a helping hand to all those who find difficulty in vocal communication with outside world.

Acknowledgements

This work would not have been possible without the help of Prof. Vijayakumar D. So, we would like to thank him for helping and guiding us throughout this research.

References

- [1] Y. Ouerhania, M. Jridia, A. Alfaloua, C. Brosseau, Optimized pre-processing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter, *Optics Communications*, Volume 289, 15 February 2013, Pages 33–44.
- [2] Poniga, F., Nedeveschi, S., Processing Dense Stereo Data Using Elevation Maps: Road Surface, Traffic Isle, and Obstacle Detection, *IEEE Transactions on Vehicular Technology*, (Volume:59, Issue: 3), March 2010.
- [3] Scharcanski, Jacob, Celebi, M. Emre, *Computer Vision Techniques for the Diagnosis of Skin Cancer*, ISBN: 978-3-642-39607-6
- [4] Le Hoang Thai, Ha Nhat Tam, Face recognition using standardized fingerprint model, *IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 3, No 7, May 2010, ISSN (Online): 1694-0784, ISSN (Print): 1694-0814.
- [5] Overview on Deaf-Blindness, National Center on Deaf-Blindness, Oct 1, 2008.
- [6] Sushmita Mitra, Senior Member, IEEE, and Tinku Acharya, Senior Member, IEEE, *IEEE Transactions on Systems, Man And Cybernetics-Part C: Applications And Reviews*, Vol. 37, No. 3, May 2007.
- [7] Jie Liu, Jigui Sun, and Shengsheng Wang, Pattern Recognition: An overview, *International Journal of Computer Science and Network Security*, VOL.6 No.6, June 2006.
- [8] W. Chen, Y. Q. Shi, and G. Xuan, Identifying computer graphics using HSV color model and statistical moments of characteristic functions, *Proceedings of IEEE International Conference on Multimedia and Expo*, Jul. 2007, pp. 1123-1126
- [9] Devi, H.K.A. 2006. Thresholding: A Pixel-Level Image Processing Methodology Preprocessing Technique for an OCR System for the Brahmi Script, *Ancient Asia* 1:161-165
- [10] T. Chai, and R. R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 7, 1247-1250, 2014, 28 Feb 2014