Air Writing Recognition and Speech Synthesis

Catharine Elizabeth Kurian[1], Chelsa Martin[2], Deepthi P J [3], Eldhose K A[4]

1,2,3 UG Student, Dept. of EEE, M.A College of Engineering, Kerala, India
4Professor, Dept. of EEE, M.A College of Engineering, Kerala, India

---***---

Abstract - *Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from face and hand gesture recognition. Users can use simple gestures to control or interact with devices without physically touching them. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques.[1] Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans than primitive text user interfaces or even GUIs(graphical user interfaces), which still limit the majority of input to keyboard and mouse.*

*Key Words***: Gesture recognition, Color tracking and detection, Character, MATLAB, Speech, Web camera**

1. INTRODUCTION

Touchless Sensing and Gesture (or Gestural) Recognition are two technologies in fast-paced growth mode because both are part of the paradigm shift in human-machine interfaces. The typical use of these technologies is to detect the motion of a person within an area covered by a sensor. To capture gestures, a camera reads the movements of the user. We use this technology to detect the letters written in air.

In the field of technology and image processing, **finger tracking** is a high-resolution technique that is employed to know the consecutive position of the fingers of the user and hence represent objects in 3D.The optical systems ,are able to capture hand motion through markers. In each hand we have markers of red and blue color. High-resolution cameras are responsible for capturing each marker and measure its positions. This will be only produced when the camera is able to see them. The visual markers are used for writing letters in air. Their motion is captured by a web camera and the data is processed using a MATLAB program. The word is finally recognized and outputted as speech through an electronic speaker.

1.1 Tracking Technique

The various steps in the tracking technique is explained below.

Background subtraction: the idea is to convolve all the images that are captured with median filtering and then these are scaled to reduce noisy pixel data.

Segmentation: a binary mask application is used to represent with a white color, the pixels that belong to the required(red) marker and to apply the black color to the foreground.

Area and centroid extraction: The area and centroid of the white region is calculated.

Comparison and character recognition: A database consisting of all the numbers(0-9),uppercase and lowercase letters are formed. The input is compared with this database and the best matching template is chosen.

This process goes on until a full word is finished. Then the entire word is spoken out using google text to speech converter.

2. BLOCK DIAGRAM

A web camera detects the color marker. Then background subtraction and noise filtering is done. The image is saved and then processed. The image is divided into lines and each line is processed. A database of all uppercase and lowercase letters and numbers is created and saved as templates pre-hand. The processed output is compared with the templates. The best matching character is outputted. So a word is formed and is then converted to speech. A block diagram which explains the above process is shown in the figure 1.
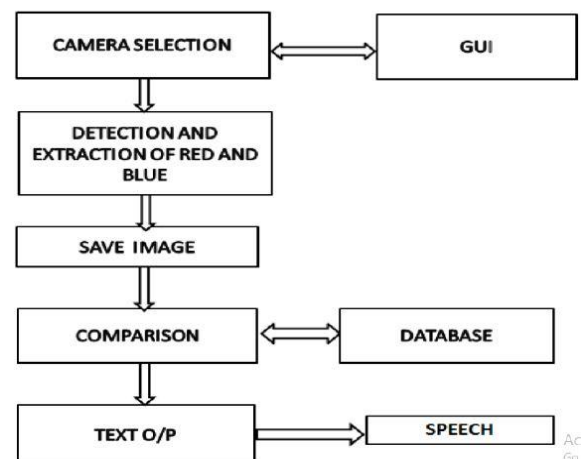

Fig -1: Basic Block Diagram of air writing recognition

Graphical user interface in MATLAB:

1. Start GUIDE by typing guide at the MATLAB prompt.

2. In the GUIDE Quick Start dialog box, select **the** Blank GUI (Default) template, and then click OK.

3. Add necessary tools from the component palette.

4. After editing then save it. When you save a layout, GUIDE creates two files, a FIG-file (information about layout) and a code file(information about MATLAB functions used)

Detection and saving of image: Image processing is done in MATLAB. The web camera detects and tracks the movement of the combination of red and blue color. If red is shown alone, the camera stops detection. after detection and extraction it is stored as a image.

Comparison with database: Template matching is done with the preprocessed characters in the database.

Text output and speech synthesis: The formed character is saved as a text file and is displayed in the GUI plain. Also the text is then converted to speech and audio is produced.

3. WORKING

The saved image is divided into lines and rows. The scanning process is done from top to bottom. Pixels of each line are summed and the text is cropped. On summation if the output is zero, then the line is deleted. Otherwise, the line is saved .It is resized to 42*24 pixel so that it matches the template which is present in the database. Now it is compared with the templates present in the database. The best matching character which gives a correlation value greater than the threshold value is taken as the output. The output character is saved to a text file and this process is repeated till the blue color alone is shown. Now for the end of the word or the sentence is recognized by zooming in the blue color.
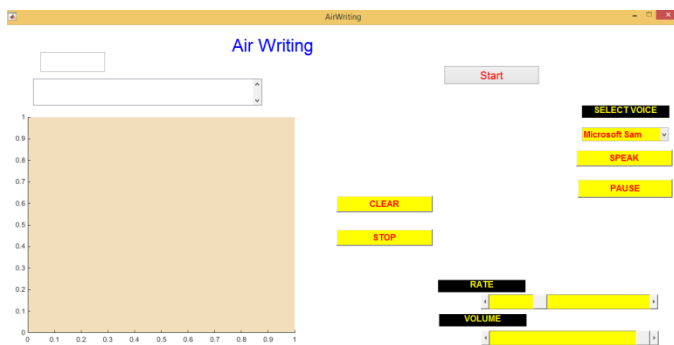
For speech synthesis there is basically there is two methods. they are based on stored sound library and MATLAB TTS function. The sound of the words are stored in the sound library. On zooming in of the blue color, the text is outputted as speech. In case speech output of the text is not present in the library then MATLAB TTS function is used.

The start button initiates the camera to continuously capture the images.

The speech output can be produced in three different built-in voices. The volume and speech output can be controlled by means of the drag bar in the output window.

The first approximation of the input is shown in the text bar at the top of the window and after thorough processing final output is shown in the text bar below it.

In case of any error while writing, it can be cleared by the clear button. Speak and pause buttons allows the user to produce the speech output in-between.

4. CONCLUSION

The conventional character inputting methods are using keyboard, digital pen, or touch screen. Here, we use a vision based character inputting system. It requires small hardware device that is camera. So it can be used for embedded applications and smart mobile phones. It acts as a reading and hearing aid for the disabled. Now, the interaction with machines is more flexible. The user does not have to wear special sensors or touch devices.

REFERENCES

[1]Kunpeng li; Xin Zhang; "A new fingertip detection and tracking algorithm and its application on writing in the air system: Image and Signal Processing(CISP)", 2014 7th International Congress on Year: 2014 Pages: 457 - 462 IEEE

[2]A.Schick, D.Morlock, C.Amma, T.Schultz, and R. Stiefelhagen,"Vision- based handwriting recognition for unrestricted text input in mid-air",in Proc. 14th ACM Int. Conf. Multimodal Interaction, 2012, pp. 217220.

[3]Lianwen Jin1, Duanduan Yang1, Li-Xin Zhen2, Jian-Cheng Huang, "A novel vision based finger writing character recognition system, Pattern Recognition", 2006. ICPR2006. 18th International Conference on human computer interaction.

[4] T. Ishikawa, Y. Horry, and T. Hoshino, "Touchless input device and gesture commands", Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers. International Conference on, pp. 205206, Jan. 2005.