Parametric method of decoding the Sign Language

Roopa S.[1], Shreesha J.[2]
[1]*Associate professor, Dept. of E & C, Siddaganga Insitute of Technology, Tumakuru*
[2]*Dept. of E & C, Siddaganga Insitute of Technology, Tumakuru*

Abstract- **Communication is a vital part of human life and is a challenge for the people who are speech/hearing impaired ones. Across the world, there are many people who are speech/hearing impaired. In their day to day life, they are facing lot of problems to communicate with others. To understand them, one has to either learn their language i.e. sign language or finger language. The system proposed in this paper aims at tackling this difficulty to some extent. In this paper, a webcam is used to recognize the hand positions and sign made. The gesture captured via webcam is converted into text/speech output which makes people understand what exactly is being conveyed. In the proposed method, the hands are first detected using skin color. Skin samples of various persons including male, female and different races are collected and the distribution of the skin pixel samples is modeled as a Gaussian in Cb-Cr color plane. This parametric method of skin/non skin classification is more accurate compared to non-parametric method. Finally, recognition of sign is performed by template matching method. The algorithm used in this paper is implemented using MATLAB.**

Index Terms-**Sign language; Gaussian model; parametric method.**

I. INTRODUCTION

Sign language is a communication skill that uses gestures as a substitute of sound to convey a message. A sign can be shown by the orientation and movement of the hands, arms or body, and facial expressions along with combining hand shapes to express fluidly a speaker's thoughts. Across the world, about 9.1 billion people are speech/hearing impaired [1]. Sign languages are commonly developed in deaf communities and often used by their interpreters, friends and family as well as people who have difficulty in hearing themselves. In hand talk implementation of a gesture recognizing glove, the gesture recognizing glove makes use of a low cost packaging material (velostat) for making piezoresistive sensors [2]. The main drawback of these works is that the user has to wear a separate hardware for recognition of gestures which makes it difficult to carry. Hand gesture recognition for human computer interaction is a novel approach based on detection of some shape based features. [3]. The setup consists of a single camera to capture the gesture formed by the user and the hand image is taken as an input to the proposed algorithm. The overall algorithm is divided into four main steps, which includes segmentation, orientation detection, feature extraction and classification. This algorithm is independent of user characteristics. It does not require any kind of training of sample data. However, the user has to show several gestures due to which time required to form a sentence is more in order to form a single word, which is undesirable [3].

In this paper, an attempt has been made to overcome the limitations of the existing methods of decoding the sign language. In this, RGB image is captured by the webcam. There are different models present for skin color segmentation like RGB, CMY, YCbCr, YCgCr, HSV etc [4]. The best color model among these has to be chosen for efficient skin color segmentation. YCbCr color model is chosen since it is better at discriminating the skin from background pixels. Image segmentation is done to separate foreground pixels from background pixels. Skin pixels are converted to white and background pixels are converted into black. OpenCV is extensively used for all simulations [5]. Gaussian model for skin segmentation [6] is both parametric and probabilistic approach that can be used to get better results. Template matching method is used for sign recognition. Sign recognition could be done on the normalized image [7].

The paper is organized as follows. Section II presents the proposed system overview for sign recognition. Section III explains the image segmentation by Gaussian method. Section IV describes the sign recognition using template matching method. Section

V represents the simulation results and section VI concludes the paper.

2.SYSTEM OVERVIEW

Conversion of Sign language to text/speech helps the speech/hearing impaired people to communicate with others. The speech/hearing impaired person gestures the sign language in front of a camera, which captures the video. The algorithm then recognizes the sign and converts it into text/speech. The block diagram used in decoding the sign language is as shown in Fig. 1.

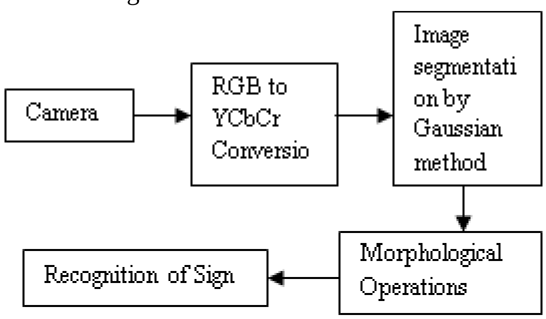

Fig 1: Block diagram of decoding the sign language.

Initially image acquisition is done wherein an image or video is acquired using a camera. The video consist of the continuous gestures shown by the person. The video processing is done in which, each frame of the video is processed for color space conversion i.e. in each frame the image is converted to YCbCr color model in order to differentiate the region of interest from the rest of the background and then further processed for image segmentation. In image segmentation, the image is converted to black and white in order to obtain the region of interest (hands and face is converted to white and rest is converted to black) using Gaussian model. The black and white image is then processed for morphological operations. Noise in the image is removed by morphological operations using dilation and erosion. After removing the noise, the final part is recognition of gesture which is done by template matching method. In this method, the original segmented black and white image will be stored in the database along with the co-ordinates of the rectangular shaped window frame defined for each sign and is compared with the output. Once the gesture is recognized, it is assigned a label and this label is displayed on the command window and is converted to speech. Skin segmentation using different color models for skin detection has many advantages as this technique is scale independent, rotation independent, and fast. By using this method, it is possible to reduce computational cost for skin detection. RGB, YCbCr and HSV color models are most popular for skin detection. Overview of these color models are as follows [5]:

RGB Color Model: This model specifies the intensity of red, green, and blue on a scale of 0 to 255, with 0 (zero) indicating the minimum intensity and 255 as the maximum intensity. YCbCr Color Model: YCbCr is a family of color spaces used in video systems. Y is the luma component and Cb and Cr are the chroma components. The YCbCr color space is a scaled and an onset version of the YUV color space. This is a brightness independent model.

HSV color Model: Hue, Saturation, Value is a color model that describes colors (hue or tint) in terms of their shade (saturation or amount of gray) and their brightness (value or luminance).

The input test image taken to analyze the color model is shown in the Fig 2. Fig 3 depicts the results obtained from different color model. In RGB and HSV color model, there is no much difference between the skin color and the background. The YCbCr color model was found to be discriminating the skin from the background in most of the test images. Hence YCbCr color model has been chosen for further processing.

Fig 2: Input Test Image

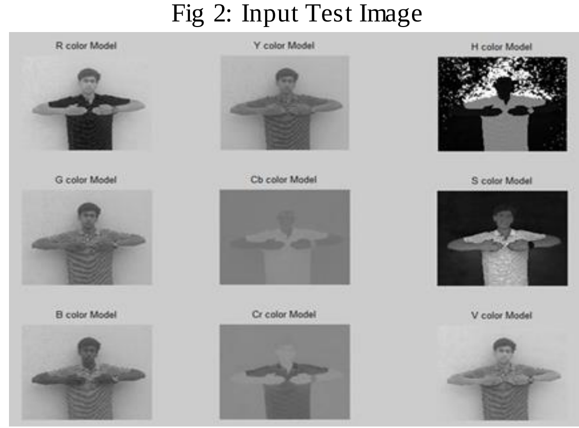

Fig 3: Results of different color models

3. IMAGE SEGMENTATION

The most popular histogram based non-parametric skin models require much storage space and their performance directly depends on the representation of the training images set. The need for more compact skin model representation for certain applications along with ability to generalize and interpolate the training data stimulates the development of parametric skin distribution models. Skin color distribution can be modeled by an elliptical Gaussian joint probability density function, defined as

$$p(c \mid skin)=\frac{1}{2\pi|\Sigma_s|^{1/2}} e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)} \quad (3.1)$$

Here c is color vector, μ_s and Σ_s are distribution parameters. A real time image with nine different skin samples is shown in Fig 4 was taken. The Cb-Cr components of this image were extracted since this model gives more accurate results for skin segmentation. The mean and covariance matrix was found for this YCbCr image which will be used later. The test image is taken and again Cb-Cr components are extracted which forms the color vector of test image. Now, each pixel of test image in color vector is compared with mean and covariance matrix. The probability of each pixel being skin is found out by using the above Eqn. (3.1). Further, the test skin samples were increased to get more accurate results as shown in Fig 5.

Fig 4: Initial nine skin samples

Fig 5: Increased test skin samples

The input image is shown in Fig. 6 and the corresponding output image of Gaussian model is shown in Fig. 7. The skin pixels in the image are segmented and a few skin like pixels in the background are also segmented.

Fig 6: Input image showing the sign "ARE"

Fig 7: Output image based on Gaussian model

The 3D histogram plot of test CbCr vector is shown in Fig. 8. The histogram plot represents the Gaussian distribution where the Cb and Cr component value of the skin sample image are concentrated in the range of 100 to 130 in case of Cb and 140 to 160 in case of Cr component.

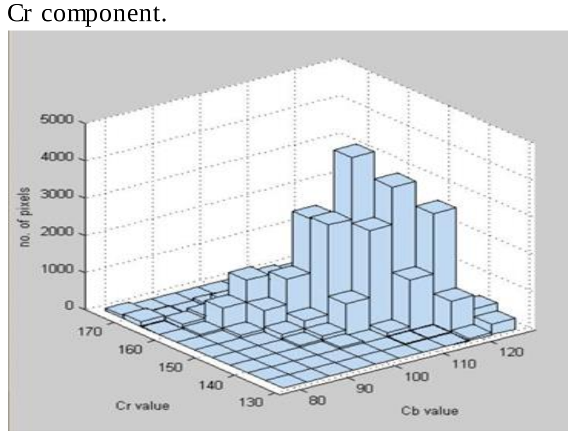

Fig 8: 3D histogram plot of CbCr test vector

4. SIGN RECOGNITION

Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, a morphological operation that is sensitive to specific shapes in the input image was constructed. The morphological operations includes

1. Dilation:- The value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to the value 1, the output pixel is set to 1.
2. Erosion:- The value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to 0, the output pixel is set to 0.

After the skin segmentation procedure using Gaussian model and morphological operations, the next step is to recognize the sign shown by a speech/hearing impaired person. This can be done using template matching technique.

1. The templates of the segmented black and white image of all signs are stored in the database and are shown in Fig 9.
2. Rectangular windows are also defined for each sign for the user's reference as shown in Fig. 10 (Rectangular windows are chosen since they are best for the hands). The database is loaded into the system.
3. The image is compared pixel by pixel with all the templates and the comparison is performed only for white pixels.
4. After comparison, total number of white pixels is calculated in each window.
5. Window having more number of ones is found and corresponding sign is assigned. The sign assigned is either displayed on the screen or it can be converted into speech.

Fig 9: Templates stored in database

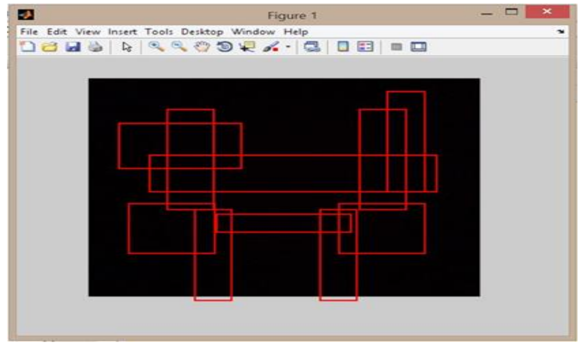

Fig 10: Rectangular windows defined for each symbol

Text To Speech Conversion

Once the gesture is recognized, the final part left out is converting it into speech. It is done by a function which uses MATLAB .NET framework interface. The Microsoft .NET framework is an integral windows component that provides a large body of pre-coded solutions to common program requirements, and manages the execution of programs written specifically for the framework. MATLAB supports the .NET framework on the windows platform only.

5. SIMULATION RESULTS

This section shows the results of image segmentation and their performance, followed by the performance of sign recognition using template matching method. Gaussian model is more accurate than all the non parametric approaches of image segmentation. Here, it takes a set of predefined skin samples so that it reduces false detection of skin. From the input image, the segmented image obtained is then converted into the corresponding text/speech output. The results of the above proposed model are shown in the following figures. Fig.11 shows the input and segmented image of a person showing sign "how". The sign is detected and the text is displayed in the left corner of command window. In the similar fashion, different gestures such as "are", "you", "where", "doing", "dad", "what" were also recognized.

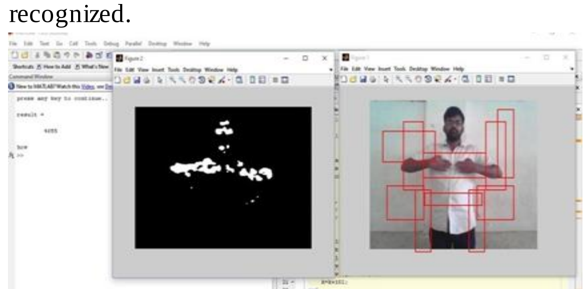

Fig 11: Segmented image and text output "how"

In Fig 12, the sign is shown for "are". The image is taken in low lighting condition. The sign is decoded and the text is displayed in the command window correctly even at low lighting condition.

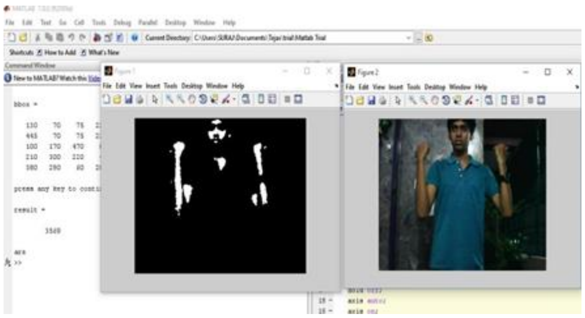

Fig 12: Segmented image and text output "are"

In Fig 13, "are" sign is shown by a person wearing skin color shirt. Skin colored shirt is also segmented and hence wrong output is displayed on the command window.

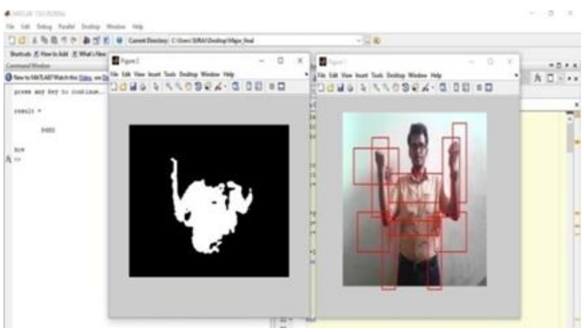

Fig 13: Case in which Gaussian model fails

6. CONCLUSION

In this project, seven different American Sign Languages such as "how", "are", "you", "where", "doing", ""dad", "what" were recognized. Probabilistic approach is used for segmentation. The better segmentation of hands is obtained when skin is modeled using Gaussian density function. Recognition of sign is done by template matching method. The output may vary due to different lighting conditions. The main criteria for this system to work properly are that the skin color and background color should be different. The best results are obtained when the background is white or black and shirt color is not skin color. The segmented image can be normalized to make it independent of distance between the webcam and the person. Sign recognition could be done with respect to the normalized image.

This type of decoding can be further improvised by making use of 'kinect' used by xbox. It also uses image processing for detecting hand gestures. Using kinect, it is possible to detect various number of signs even with similar signs. This has a drawback with less number of signs. Hence, it is proposed to use kinect to increase the number of words.

REFERENCES

[1] Pathan Y.M.,.Waghmare S.R, Patil P.K., "Sign Language Recognition for Deaf and Dumb People", International Journal of Engineering and Computer Science, ISSN:2319-7242, Volume 4, Issue 3, pp. 10872-10874, March 2015.

[2] Preetham C., Ramakrishnan G., Kumar S., Tamse., Krishnapura N., "Hand Talk Implementation of a Gesture Recognizing Glove", India Educators' Conference (TI-IEC), Texas Instruments, pp. 328-331, April 2013.

[3] Panwar M., Mehra P.S., "Hand gesture recognition for human computer interaction", International Conference on Image Information Processing (ICIIP), pp. 1-7, Nov 2011.

[4] Mrunmayee V. Daithankar, Kailash J. Karande, Avinash D. Rarale. "Analysis of Skin Color Models for Face Detection", International Conference on Communication and Signal Processing, pp. 533-537, April 2014.

[5] Ivan Culjak, David Abram, Tomislav Pribanic, Hrvoje Dzapo, Mario Cifrek, "A brief introduction to OpenCV", MIPRO, pp 1725-1730 , Opatija, Croatia, May 2012

[6] Vladimir V., Vassili S., Alla A.,"A Survey on Pixel Based Skin colour Detection Techniques", in proceedings of the GraphiCon, pp. 85-92, July 2003.

[7] Lingxi Xie, Qi Tian, Bo Zhang, "Simple Techniques Make Sense: Feature Pooling and Normalization for Image Classification", IEEE Transactions on Circuits and Systems for Video Technology, Volume 26, Issue 7, pp. 1251-1264, July 2016.