ISSN:2231-0711
www.ijcset.net

SignPro-An Application Suite for Deaf and Dumb

Ashish Sethi[#1], Hemanth S[#2],Kuldeep Kumar[#3],Bhaskara Rao N[#4],Krishnan R[#5]

Dayananda Sagar College of Engineering,
Visvesvaraya Technological University, Bangalore, India.
[1]ashishsethi1212@gmail.com
[2]hemanths90@yahoo.com
[3]kuldeep17290@gmail.com
[4]bhaskararao.nadahalli@gmail.com
[5]krishnanr1234@gmail.com

***Abstract*- This application helps the deaf and dumb person to communicate with the rest of the world using sign language. Suitable existing methods are integrated in this application. The key feature in this system is the real time gesture to text conversion. The processing steps include: gesture extraction, gesture matching and conversion to speech. Gesture extraction involves use of various image processing techniques such as histogram matching, bounding box computation, skin colour segmentation and region growing. Techniques applicable for Gesture matching include feature point matching and correlation based matching. We have come up four different approaches based on the methods used for gesture extraction and matching. A Comparative study of these approaches is also carried out to rank them based on time efficiency and accuracy. The other features in the application include voicing out of text and text to gesture conversion.**

Keywords—**Bounding box approach, correlation approach, histogram matching, point matching algorithm, region growing approach, skin color segmentation and text to speech conversion application**

I. INTRODUCTION

A deaf and dumb person uses sign language for communication. So, the person to whom this communication is intended needs to know the sign language in order to understand it. But many people may not be familiar with the sign language and as a result deaf and dumb people find it hard to communicate with others and gel with the society. This paper tries to provide a solution for this problem in which it converts a sequence of hand gestures into text which will be ultimately rendered as spoken words. Each hand gesture corresponds to a character in American Sign Language(ASL). On the other hand when a common person types in a sentence to the application, it converts it into a sequences of hand gestures for the deaf and dumb to see and understand. Human Computer Interaction (HCI) and image processing are related areas of research which help in the solution to this problem.

This paper proposes a set of image processing based solutions for this problem namely: skin colour segmentation with point matching, color segmentation with correlation based matching, region growing approach with point matching, region growing with correlation based matching. A comparison of these approaches is carried out on key factors like time performance and accuracy. The best among these approaches is identified and used in the implementation.

The next section presents related work on this area. Section 3 details our novel algorithm. Section 4 discusses our experimental results followed by the conclusion.

II. RELATED WORK

In recent years there has been a lot of research on hand gesture recognition. Several techniques have been reported on gesture recognition which includes skin segmentation using color pixel classification [1], region growing by exemplar-based hand segmentation under complex background [2], Parametric Hidden Markov models for gesture recognition [3], statistical database comparison method [4], accelerometer-based gesture recognition system [5], orientation histograms for gesture recognition [6], Finger Detection for Sign Language Recognition[7] etc. Most of the gesture recognition systems use special devices like hand glove[11]. The gloves get connected to the computers using a lot of cables. So these devices are cumbersome and expensive. In order to overcome these difficulties, alternatively vision-based approaches involving camera and image processing for recognizing gestures are being explored.

There has been previous work on use of features for finger detection. Some common features extracted include hand silhouettes[9],[10], contours[11], key points distributed along fingertips & joints[12][13][14]. There has also been reported work where finger detection has been accomplished via color segmentation, real-time hidden Markov model-based systems[15], computational model of intelligibility for ASL (CIM-ASL)[16], enhanced level building (eLB) algorithm [8] and contour extraction**[**18**]**. But colour based segmentation techniques requires fine-tuning every time the system switches to a new user as the color complexion varies from person to person. With the limitations posed by the schemes discussed above there is scope to devise a more efficient and robust technique.

III. PROPOSED METHODOLOGY

In this section we present a robust and efficient technique for gesture recognition. As shown in fig. 1, our method has four main phases of processing**:**

- Pre-processing mechanism
- Gesture extraction
- Gesture matching
- Conversion of the text to voice

For gesture extraction phase, two methods are used, namely:

- Skin color segmentation
- Region Growing

For gesture matching phase, two methods are used, namely:

- Feature point matching using SIFT
- Correlation matching

Methods for gesture extraction and matching are integrated suitably leading to following four approaches, namely:

Approach A: Skin color segmentation with Feature point matching using SIFT
Approach B: Region Growing with Feature point matching using SIFT
Approach C: Skin color segmentation with Correlation matching
Approach D: Region Growing with Correlation matching

In addition we have presented text to gesture conversion application.
The Region of Interest (ROI) is identified by applying skin segmentation, region growing exemplar based technique and contour segmentation. Since we cannot handle the complex background using contour based approaches we use skin segmentation and region growing for obtaining ROI.
Correlation based recognition approach is scale and rotation variant and will not work for rotated or size scaled image. We handle the scale variance by normalizing the size of captured image before processing. In order to address the rotation variance, we take 9 pictures of a single gesture in different orientation with a difference of 10 degrees between consecutive orientation. On the other hand point based methods like SIFT[17] is scale and rotation invariant and hence there is no additional processing required for making it view invariant.

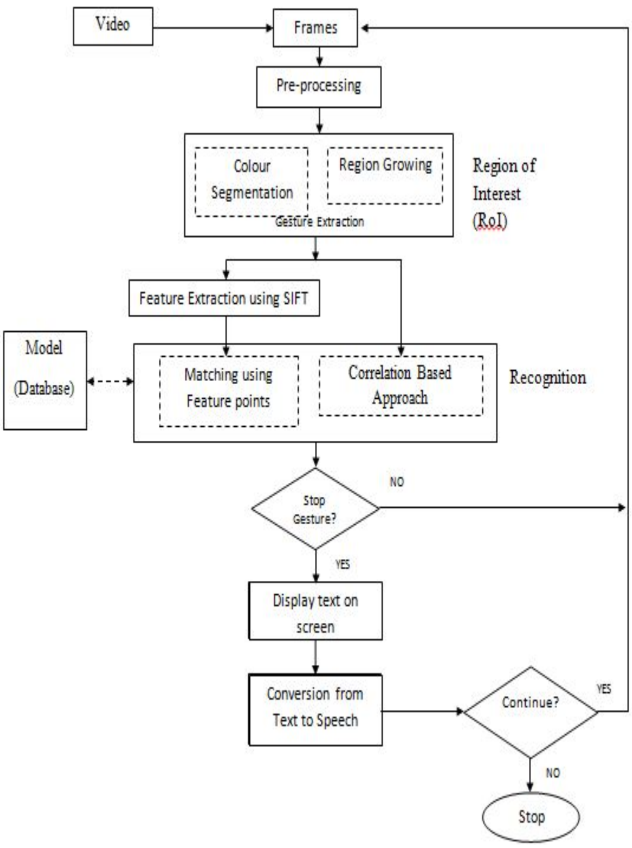

Fig1: Flow chart of Gesture to Text conversion

A. *Pre-processing mechanism*

On each of the input image, histogram equalization is performed. Here we have a standard reference image with good brightness, illumination and intensity factors. All the input images will be equalized to the reference image. So, all the input images will have the same amount of brightness and other factors.
Histogram equalisation enhances the contrast of images as shown in fig. by transforming the values in an intensity image, or the values in the colormap of an indexed image, so that the histogram of the output image approximately matches a specified histogram.

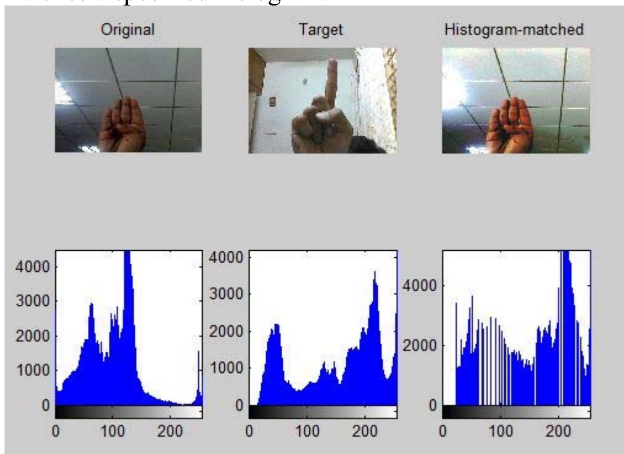

Fig. 2. Histogram equalization

B. *Gesture extraction*

Two different methods for gesture extraction are used namely: skin color segmentation and region growing approach

1) Skin color segmentation:

Skin segmentation focuses on producing the skin map of a given image.
Following are the steps involved in this method:

1. Identify skin patches

Fig. 3: Image with skin patches

2. Calculation of area of several patches and retain the patch with largest area.

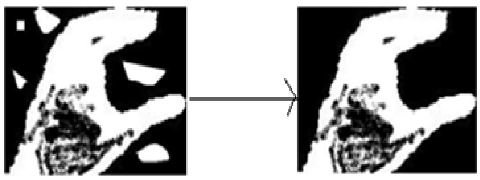

Fig.4: Obtaining skin patch with maximum area

3. Filling of the largest pattern to remove irregularities.

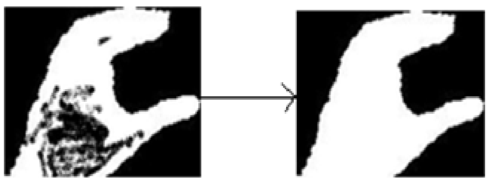

Fig.5: Results after filling irregularities

2) Region growing approach:

In this approach we use connected component algorithm[2] for identifying the hand region . In order to make region growing computationally efficient, a block-based updating process is presented. This method works well even on input image with noise, shadow and occlusion.

First, given an input image, the system selects the centre pixel to be the seed pixel for the hand region Ω to be grown. The remaining region is defined as the source region ϕ(ϕ=1- Ω). Next, we set the size of the template window Ψ as 3*3 pixels. Once these parameters are defined, the region-growing proceeds automatically. The four main steps of our hand segmentation algorithm are as follows:

Step 1: Computing patch priorities.

Given a patch p ψ centred at the point p for some $p \epsilon \partial\Omega$, where $\partial\Omega$ denotes the boundary of Ω. This paper defines its priority $P(p)$ as the product of the three terms:

$$P(p)= D(p)C(p)M(p)$$

Where $D(p)$, $C(p)$ and $M(p)$ respectively depict the data term , the confidence term and the absolute value of mean difference term.

Step 2: Updating Region Growing

Once all priorities on the region front have been computed, the patch Ψ_p with the least priority is found. Region growing data is extracted from the source region ϕ. After updating region growing, the hand segmentation algorithm updates the front related to patch Ψ. Then update the data term $D(p)$, which is near to the patch Ψ

Step 3: Updating Confidence Values

After the patch Ψ_q has been updated with new pixel values, the confidence $C(p)$ is updated as follows: $C(p)=C(q)$, $\forall p \ \epsilon \ \Psi_q \cap \Omega$. The rule is simple, but it measures the relative confidence of patches on the region front without specific image parameters.

Step 4: Updating the Absolute Value of Mean Difference

After updating the region growing mean, then the algorithm updates the absolute value of mean difference on the region front

Fig6. Hand segmentation under complex background

C. Gesture matching

Two approaches are used here: point matching algorithm and correlation based approach.

1) Point matching algorithm:

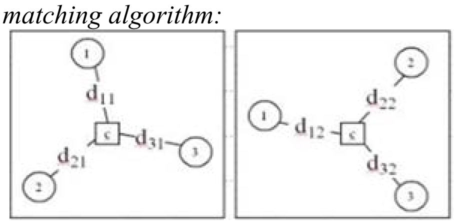

Fig. 7 Input image versus database image

Step 1: Each input image must be compared against the set of all database images.

Step 2: Find the SIFT keypoints in the input image using distance threshold. The distance threshold retains only those matches in which the ratio of vector angles from the nearest to the second nearest key-point is less than the threshold we specify.

$d_{T1}= \sum_{i=1}^{M} d_{i1}$

$d_{T2}= \sum_{i=1}^{M} d_{i2}$

Step 3: For each of the found descriptor in the first image, select its match in the second image.

Step 4: Calculate the distance between the matched key-point with the centre point of the image.

Step 5: Calculate the distance ratio.

Ratios1=[d_{11}/d_{T1} d_{21}/d_{T1} d_{31}/d_{T1}]

Ratios2=[d_{12}/d_{T2} d_{22}/d_{T2} d_{32}/d_{T2}]

Step 6: Mark the distances which are below the algorithm's threshold.

Distance Mask= abs[Ratios1-Ratios2]

Valid Points= sum(Distance Mask)

Validity Ratio= (# of Valid Points)/(# of Matched Points)

Step 7: Find the image which has maximum of matches and obtain the corresponding alphabet.

2) Correlation based approach

Here we find the correlation co-efficient between the database image and the input image so that the best match is obtained.

The correlation coefficient is computed between A and B, where A and B are matrices or vectors of the same size.

Correlation Co-efficient :

Correlation(r) =[$N\Sigma XY - (\Sigma X)(\Sigma Y)$ / Sqrt([$N\Sigma X^2 - (\Sigma X)^2$][$N\Sigma Y^2 - (\Sigma Y)^2$])]

where

N = Number of values or elements

X = First Score

Y = Second Score

ΣXY = Sum of the product of first and Second Scores

ΣX = Sum of First Scores

ΣY = Sum of Second Scores

ΣX^2 = Sum of square First Scores

ΣY^2 = Sum of square Second Scores

D. Text to gesture conversion

In this part user enters the sentence which is extracted for its characters. Each of these extracted characters are then further matched with the database gesture image stored for that character. And finally the user obtains sequence of gestures for the sentence in a single window

HOW ARE THEY

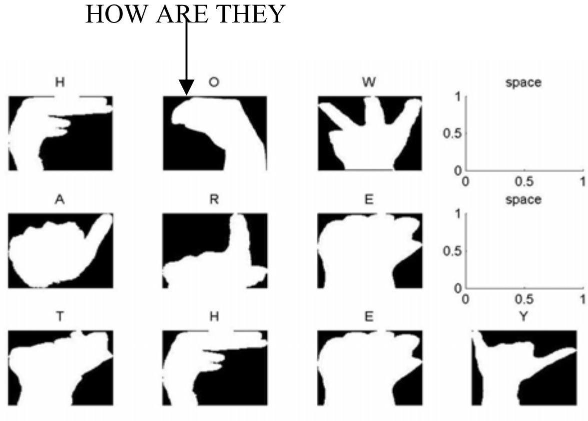

Fig8. Text to Gesture conversion

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Testing is performed on 600 test samples with different hand gestures in a complex background[19]. The dataset contains the different hand gestures of all the alphabets. The results for time efficiency and accuracy as shown below:

Table 1: Time performance (seconds)

Time Performance (seconds)

Gesture Matching \ Gesture Extraction	Skin color segmentation	Region Growing
Feature point matching through SIFT	A 26.045	B 35.364
Correlation matching	C 2.667	D 12.729

Table 2: Accuracy (%)

Accuracy (%)

Gesture Matching \ Gesture Extraction	Skin color segmentation	Region Growing
Feature point matching through SIFT	A 84.15	B 87.88
Correlation matching	C 73.57	D 76.34

Based on the experimental results it is observed that approach C which consists of Skin color segmentation with Correlation matching gives the best time efficiency. Experimental results also show that approach B which consists of Region Growing with Feature point matching using SIFT gives the best accuracy of matching.

V. CONCLUSIONS

We have designed and implemented four approaches for gesture to text conversion. Therefore, with the experimental results obtained for the four above approaches, the approach of skin color segmentation method for gesture extraction and correlation approach for the matching is analysed to be the most time efficient approach for the application in terms of time complexity. The approach consisting of Region Growing with Feature point matching using SIFT gives more accurate matching. We have also designed and implemented text to gesture conversion feature where a sentence is converted into a sequence of hand gestures and displayed as output.

The future enhancement for this paper is to port it to android platform running on mobile for its broader use. Machine learning approaches can also be used to enhance efficiency and experimental accuracy.

REFERENCES

[1] Son Lam Phung, Member, IEEE, Abdesselam Bouzerdoum, Sr. Member, IEEE, and Douglas Chai, Sr. Member, IEEE, "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison", IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 1, january 2005

[2] Ma De-yi, Chen Yi-min, Li Qi-ming, Huang Chen, Xu Sheng, "Region Growing by Exemplar-Based Hand Segmentation under Complex Backgrounds"

[3] Andrew D. Wilson, Student Member, IEEE Computer Society, and Aaron F. Bobick, Member, IEEE Computer Society, “Parametric Hidden Markov Models for Gesture Recognition”, IEEE transactions on pattern analysis and machine intelligence, vol. 21, no. 9, september 1999

[4] Chance M. Glenn, Divya Mandloi, Kanthi Sarella, and Muhammed Lonon, "An Image Processing Technique for the Translation of ASL Finger-Spelling to Digital Audio or Text"

[5] Ahmad Akl, Student Member, IEEE, Chen Feng, Student member, IEEE, and Shahrokh Valaee, Senior Member, IEEE, “A Novel Accelerometer-Based Gesture Recognition System”, IEEE transactions on signal processing, Vol.59, No.12, December 2011

[6] William T Freeman, Michal Roth, “Orientation Histogram for Hand Gesture Recognition”, December 1994, http://www.merl.com

[7] Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V Pujari, "Finger Detection for Sign Language Recognition", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong

[8] Ruiduo Yang, Sudeep Sarkar, Senior Member, IEEE, and Barbara Loeding, "Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming", IEEE Transactions on Pattern Analysis and Machine Intelligence,, vol. 32, no. 3, march 2010

[9] N. Shimada et al., “Hand gesture estimation and model refinement using monocular camera—Ambiguity limitation by inequality constraints,” in Proc. Face and Gesture Recognition, pp. 268-273, 1998.

[10] M. Sonka, V. Hlavac, and R. Boyle, Image processing, Analysis, and Machine Vision, Brooks/Cole Publishing Company, 1999.

[11] T. Starner and A. Pentland, “Real-time American Sign Language recognition from video using hidden Markov models,” in Proc.of International Symposium on Computer Vision, pp. 265–270, 21-23 Nov.1995.

[12] J. Lee and T. Kunii, “Model-based analysis of hand posture,” IEEE Comput. Graph. Appl., vol. 15, no. 5, pp. 77-86, Sept. 1995.

[13] H. Fillbrandt, S. Akyol, K. F. Kraiss, “ Extraction of 3D Hand Shape and Posture from Image Sequences for Sign Language Recognition.” IEEE International Workshop on Analysis and Modeling of Faces and Gestures, vol. 17, pp. 181-186, October 2003.

[14] T. Starner, A. Pentland , J. Weaver, “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20 issue 12, pp. 1371-1375, December 1998.

[15]Thad Starner, Student Member, IEEE, Joshua Weaver, and Alex Pentland, Member, IEEE Computer Society, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, December 1998

[16] Frank M. Ciaramello, Student Member, IEEE, and Sheila S. Hemami, Fellow, IEEE, A Computational Intelligibility Model for Assessment and Compression of American Sign Language Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, November 2011

[17] Faraj Alhwarin, Danijela Ristić–Durrant and Axel Gräser, Institute of Automation, University of Bremen, Otto-Hahn-Alle NW1, D-28359 Bremen, Germany, {alhwarin, ristic, ag}@iat.uni-bremen.de, “Speeded Up Image Matching Using Split And Extended Sift Features”

[18] Lalit Gupta and Suwei Ma, “Gesture-Based Interaction and Communication: Automated Classification of Hand Gesture Contours”

[19] http://www.buzzle.com/articles/sign-language-for-kids.html