Double Handed Indian Sign Language to Speech and Text

Kusumika Krori Dutta[1], Satheesh Kumar Raju K[2], Anil Kumar G S[3], Sunny Arokia Swamy B[4]
[1, 2, 3, 4]Department of Electrical and Electronics Engineering
M.S.Ramaiah Institute of Technology, Bengaluru-560054
[1]kusumika@msrit.edu, [2]satheeshkumarraju21@gmail.com, [3]anilk6871@gmail.com, [4]sunny23june@gmail.com

***Abstract*—Sign Language is the main communicating tool for hearing impaired and mute people, and to ensure an independent life for them, the automatic interpretation of sign language is an extensive research area. With the use of image processing and artificial intelligence, many techniques and algorithms have been developed in this area. Every sign language recognition system is trained for recognizing the signs and converting them into required pattern. The proposed system aim to provide speech to speechless, in this paper the double handed Indian Sign Language is captured as a series of images and it's processed with the help of MATLAB and then it's converted to speech and text.**

Keywords—Indian Sign Language; Deaf and Mute; Translator; Image Processing; MATLAB.

I. INTRODUCTION

Communication is the act of transferring information from one person to another. Communication plays a vital role, without this existence of human life on earth cannot be imagined. In general people use oral or written communication to communicate but hearing impaired and mute people are not blessed with oral communicating ability and that makes them extremely isolated and lonely. The invention of sign language enabled them to share their feelings to the people having the knowledge of sign language

Sign language is a system of communication using visual gestures and signs, as used by deaf and mute people. It is the well-structured code gesture where every gesture has a specific meaning assigned to it. There are 143 existing different sign languages (types with dialects) across the world, mainly American Sign Language, British Sign Language, French Sign Language, Japanese Sign Language, and Indian Sign Language [1]. Even some nations have more than one sign language. Sign Language is the only means of communication for deaf and mute people mainly illiterates. But to communicate with normal people without a human translator still is a challenge as most of the normal people in the society are not interested to learn this sign language. his causes isolation of the mute and hearing impaired people. But with the advancement of technologies it is possible to overcome from the challenge and bridge the communication gap. The machine can be programmed to recognise the gestures and interpret them to speech and text.

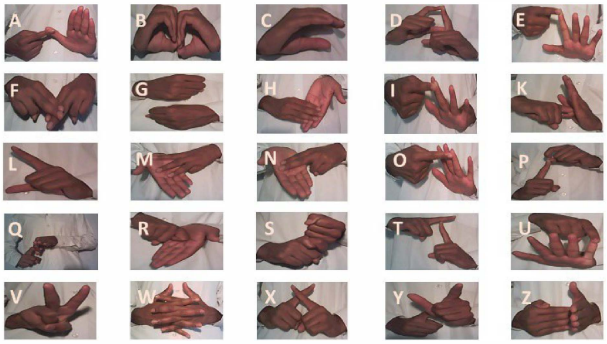

Figure. 1. Double handed Indian Sign Language

In India about 6.3% of its population are deaf and mute. In a developing country like India, where literacy rate is less it is difficult for ambitious illiterate people to mingle and interact with the fellow beings. Indian Sign Language can be communicated using Single hand and Double hand, figure1 shows the different gestures of double handed Indian sign language

The proposed system provides voice to the deaf and mute people and promising them an independent life without any help of human translator. The system is trained with double handed sign language by using a minimum eigenvalue algorithm. Here Logitech web camera is used for image acquisition and processing is performed in MATLAB. The corresponding output is obtained after extracting Shi-Thomasi good features.

II. BACKGROUND

Though sign language is a combination of hand gestures, facial expressions and body language, hand gestures provides majority of the information and hence majority of the research is going on decoding the hand gestures. This section deals with the past and ongoing work in the field of gesture recognition for various applications. Most work in this research field tries to elude the problems of illumination and cluster background by using markers and data gloves.

The conversion of sign to text is proposed [2]. Authors in

978-1-5090-0148-4/15/$31.00© 2015 IEEE

this paper proposed a method on conversion of signs to text. In this paper, two different methodologies have been implemented. One methodology is based on statistical methods and another is a centroid algorithm. In statistical method, single gesture with 10 different images was processed with proper threshold was fixed [2]. In Centroid algorithm, number of active fingers were calculated after forming two concentric windows. The main drawback of these algorithms are they are position and background dependant which restricts its usage.

Praveenkumar S, et.alimplemented a method which recognizes gestures and translate into speech using PIC microcontroller and data gloves [3]. The data glove is fitted with flex sensors along the length of each finger and the thumb. The PIC microcontroller provides the corresponding voice output using APR9600.

ArtiThorat, et.al published paper which deals with gesture recognition using SIFT algorithm. The Feature vectors obtained from this algorithm are distinctive and invariant to any scaling, rotation or translation of the image [4]. The relative positions between them in the original scene shouldn't change from one image to another which is a drawback in real time situation.

Divya S, et.al proposed a SLR system which involves mostly two handed movement gestures without any non-manual markers. In this algorithm only regions of interest are the two hands and the face region. Here input video is taken from digital camera and converted into frames. Input video is converted into frames and Skin color based segmentation is employed [5]. Later Kalman filter based tracking and sift based gesture recognition is used to obtain the output. The system requires a constant background and also uniform lighting conditions.

Y Madhuri, et.al proposed a sys-tem which presents a mobile based sign language translation device for automatic translation of Indian sign language into speech in English by using LABVIEW software. This sys-tem is able to recognize one handed sign representation of alphabets and numbers [6]. The results are found to be highly consistent and accurate. But the results are light sensitive and background dependent.

Many researchers were proposed for sign language recognition but all the algorithms and methodologies proposed have some drawbacks. Thus we propose a system which converts Indian sign language to text and speech. Various difficulties faced by researches have been tried to minimize with our approach. The Proposed system is based on minimum eigenvalue algorithm. It employs Shi-Tomasi corner detector. The feature detection method used here is Tomasi's good features to track [7].

III. MINIMUM EIGEN VALUE ALGORITHM

Traditionally there exist three types of features: edges, corners and blobs. In our application we are mainly interested in corners because they are resistant to the aperture problem and are well suited to be tracked. We used the Shi-Tomasi method also known as Minimum Eigen Value algorithm to detect feature points. The Shi-Tomasi detector is a further development of the Harris corner detector. The slight divergence between those two detectors lies in the scoring function that is used to decide for each pixel of an image to mark it as a corner [8]. In order to extract good features for tracking, the patches in the same neighborhood are compared by employing auto-correlation. The corner points in the image which are a scalar object consisting of feature points detected in a 2-D grayscale input image. The corner detection takes place by looking at the intensity values within a small window. The sum of scored differences (SSD) are calculated and if SSD is low then patches match well. Suppose if an Image patch W (windowing function) at x is shifted by a small amount x. Then, the sum-squared difference at x is given by equation (1),

$$E(x) = \sum_{xeW}[I(x) - I(x + \Delta x)]^2 \qquad (1)$$

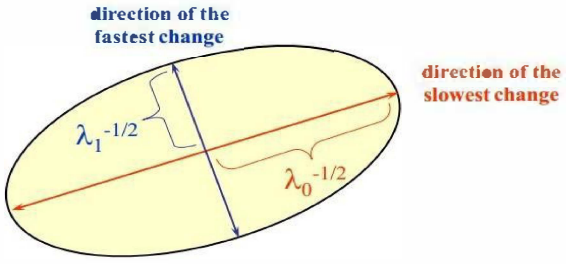

Figure. 2. Uncertainty ellipse corresponding to an eigenvalue analysis of the auto-correlation matrix A

Shi and Tomasi considered weighted auto-correlation is given by equation (2).

$$E(x) = \sum_{xeW} W(x)[I(x) - I(x + \Delta x)]^2 \qquad (2)$$

On applying Taylors series auto-correlation matrix A is shown in equation (3).

$$A = W * \begin{bmatrix} {I_x}^2 & I_x I_y \\ I_x I_y & {I_y}^2 \end{bmatrix} \qquad (3)$$

A is a 2x2 matrix which depicts the existence of scalar quantities λ 1, λ 2 which are the eigenvalues. Variations in the eigen values is shown in Figure 2.

A score is calculated for each pixel, and if the score is above a certain value, the pixel is marked as a corner. The score is calculated using two eigenvalues. The function shown below in equation (4) manipulates them, and give back a score.

$$R = \min(\lambda_1, \lambda_2) \qquad (4)$$

Classification of image points using eigenvalues:

- If both $_i$ are small, then feature does not vary much in any direction which indicates uniform region
- If the larger eigenvalue $_1$ $_1$, then the feature varies mainly in the direction of v1 which represents edge.
- If both eigenvalues are large then feature varies significantly in both directions

IV. IMPLEMENTATION

The block diagram given below explains the overall implementation of the project.

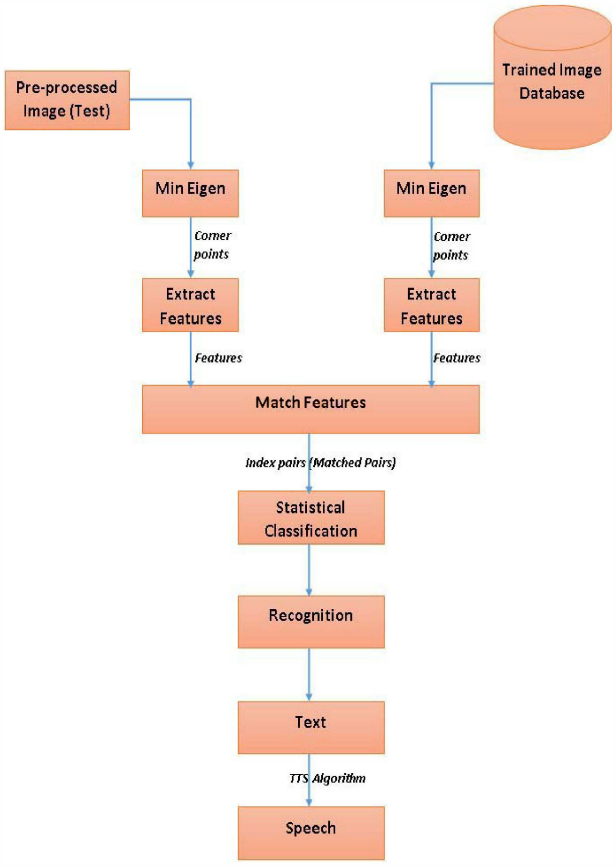

Figure. 3. Block Diagram of Min Eigen Value Algorithm

Five Images of each alphabet are taken and Min Eigen Value algorithm is applied as shown in figure 3, and interesting points are extracted. The Extracted features are stored instead of directly storing the image. This has a main advantage that it takes very less space and computational time will also be less. By this way the system is trained with image features. During real time acquisition. The image is captured by Logitech Web camera of 5MP. The image is pre-processed to reduce noise and artifacts and then its Min Eigen Value features are extracted.

Now the matching is done between extracted features of real time acquired image and that of features stored in data base. The Statistical calculation is done for the matched pairs and then its recognized and equivalent text is being displayed. The following table gives the comparison of features matching with database.

Table 1. Comparison of features matching with database

Alphabets						
A	**284**	13	37	25	16	26
B	8	**332**	16	13	13	13
C	4	43	19	35	18	16
D	31	21	12	12	39	24
E	90	30	30	43	32	32
F	7	16	**392**	22	47	39
G	13	9	34	31	49	43
H	12	23	69	32	45	39
I	24	5	23	26	23	23
K	16	6	19	31	32	22
L	16	27	32	**387**	39	15
M	10	14	42	26	41	40
N	4	18	41	12	35	33
O	21	10	20	31	53	36
P	17	7	16	22	43	41
Q	16	15	34	38	33	50
R	12	10	32	10	29	22
S	7	9	24	17	**400**	36
T	20	17	34	33	38	39
U	4	16	19	23	53	12
V	2	15	16	57	20	26
W	20	7	42	34	50	68
X	24	12	40	38	28	92
Y	14	12	33	16	35	**597**
Z	11	5	12	27	25	34

The above table shows the number of points a gesture is matching with the features of stored database. For example, the first column gesture which is showed is Indian Sign language 'A', and it's clearly observed that there are 284 points matched with A database. If any alphabet is given as input then the corresponding alphabet would have highest number of matching point which enables us to display the equivalent text result.

V. TEXT TO SPEECH SYNTHESIS

The Output Text which is obtained after Image and Word processing is further converted to Speech using Text to Speech Synthesis. The Block diagram shown in figure 4 explains the Text to Speech Synthesis.

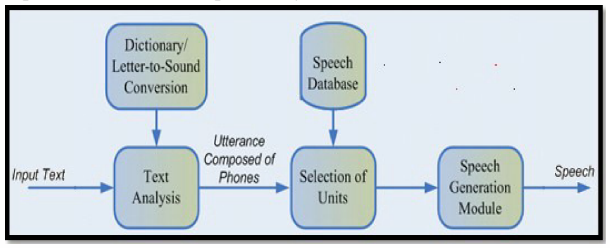

Figure. 4. Block Diagram of Text to Speech Synthesis

The Text is given as Input. The Text is analysed and each and every word is being searched in the Letter to Sound Dictionary. Then its equivalent which is been stored in the dictionary is being played in sequence of the text generated. This text was translated to speech form.

VI. RESULTS

The tabular column shown in the Fig 3 gives the count of matched alphabet against the database images. Here we can observe for test image of alphabet A the count against A in the database is highest (284) for other alphabets. Similarly the count for alphabets B, F, L, S, and Y against database images is calculated and displayed.

The feature points of trained and test image for alphabet A is shown in figure 4 and figure 5 respectively. The matched points of the database image and test image is shown in figure 6. By matching the feature points of test image as shown in figure 7, and image in database (125 images) the corresponding output was displayed as text and later translated to speech using COM server in MATLAB.

The experiment was carried out with bare hands, thus limitations by using data gloves is overcome. The results obtained were background and person independent. This algorithm was also implemented for single handed Indian Sign Language and text output was obtained.

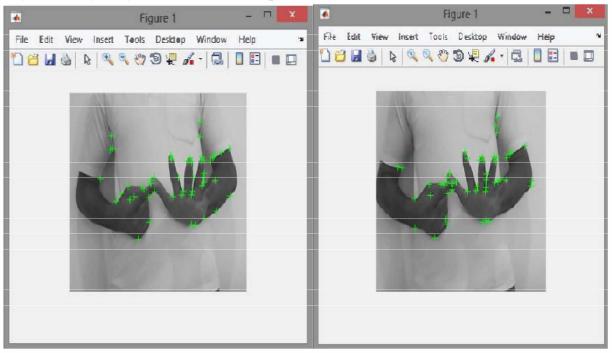

Figure. 5. Trained image Figure. 6. Test image

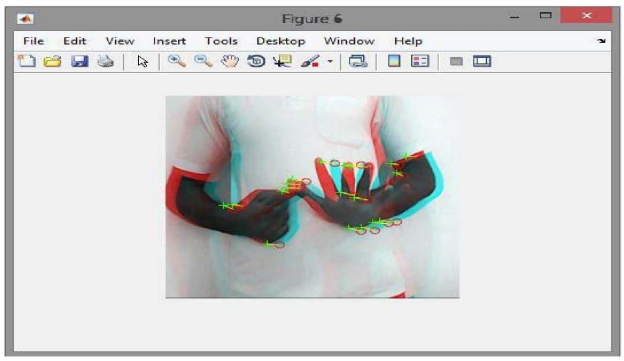

Figure. 7. Matched features

VII. CONCLUSION

Thus system was developed to translate the double handed Indian Sign Language into both text and speech This system helps and aid the hearing impaired and mute people to live independently. It develops confidence and will power to share their emotions, thoughts, ideas and difficulties with the normal people in the society. This eliminates the gaps among the people and achieve better society

This system is also useful to all the people in the society gets as through this speech translation is possible, system output can be obtained in different languages.

.

REFERENCES

[1] M.Jerin Jose V.Priyadharshni et.al, "Indian Sign Language (ISL) Translation System For Sign Language Learning,"IJIRD, vol. 2, May,2013, pp. 358–365.

[2] KonduruSatheesh Kumar Raju Sunny ArokiaSwamy B Anil Kumar G S KusumikaKrori Dutta, "Indian Sign Language to Speech," IJAER, vol.8 issue VI, December, 2014, pp. 6–12.

[3] Praveenkumar S HavalagiShruthiUrfNivedita, "The Amazing Data Gloves that give voice to the voiceless," IJAET, vol.6, issue 1,March,2013,pp. 471–480.

[4] ArtiThorat, VarshaSatpute, AratiNehe, TejashriAtre, Yogesh R Nargoje, "Indian Sign Language Recognition System for Deaf People," IJARCCE, vol.3, issue 3,March,2014.

[5] Divya S, Kiruthika, S Nivin Anton A L and Padmavahi S, 'Segmentation, Tracking And Feature Extraction For Indian SignLanguage Recognition," IJCSA, vol.4, no.2, April,2014, pp. 57– 72.

[6] YellapuMadhuri, Anitha G, Anburajan M1, "Vision-Based Sign Language Translation Device," International Conference on Information Communication and Embedded Systems (ICICES), IEEE, 2013, pp. 565– 568.

[7] Shi, J., and C. Tomasi, "Good Features to Track," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1994, pp. 593600

[8] Rajam P S, Balakrishnan G,"Real Time Indian Sign Language Recognition System to aid deaf-dumb people", Communication Technology (ICCT), 2011 IEEE 13th International Conference, pp. 737-742.

[9] M.W. Kadous, "Machine recognition of Australian signs using power gloves: Toward large-lexicon recognition of sign language," Proc. Workshop Integration Gesture Language Speech, pp. 165-174, 1996.

[10] Geetha M, Manjusha U. C, "*A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation*" International Journal on Computer Science and Engineering (IJCSE), March 2012.

[11] A. Park, S. Yun, J. Kim, S. Min, and K. Jung. "*Real time vision-based Korean finger spelling recognition system*". Proc. World Academy of SET, 34:293298, 2008.

[12] R. Rokade, D. Doye, and M. Kokare, "*Hand Gesture Recognition by Thinning Method*" International Conference on Digital Image Processing (2009).

[13] Nadia R. Albelwi, Yasser M. Alginahi, "*Real-Time Arabic Sign Language (ArSL) Recognition*" International Conference on Communications and Information Technology 2012.

[14] Rekha, K.; Latha, B. "*Mobile Translation System from Speech Language to Hand Motion Language*", Intelligent Computing Applications (ICICA), 2014 International Conference on, On page(s): 411 – 415.