

English Letters Finger Spelling Sign Language Recognition System

D.Karthikeyan^{#1}, Mrs.G.Muthulakshmi^{#2}

[#]M.S.University,Tirunelveli,India

¹PG Scholar,

²Assistant Professor,

Abstract - In this paper “English Letters Finger Spelling Sign Language Recognition System” implemented an image processing technique and Neural Network, designed for recognizing the sign language for deaf-dump person. The main objective of this paper is to convert the Fingerspelling image English [26 letters] into English speech for deaf-dump person. It will recognize and identify the finger spelling images which are presented in the English speech. 20 highest DCT coefficients are extracted from each finger spelling and are given as input to SVM model. Support Vector Machine (SVM) is used to recognize the sign (finger spelling) and play and display corresponding English speech and letters. In SVM have three kernels types Polynomial, Gaussian and Sigmoidal it's applied for Finger spelling recognition into speech.

Keywords : *Fingerspelling, SVM, DCT, SRM, Zigzag*

I. INTRODUCTION

In the 21st century field of science and technology has reached such a level that people are expecting more comfortable and useful things, which can make their lives easier. Now days, homes with voice recognition built in with the sense of gestures have already been conceived. There are video games are in the market, which can be played with real time gestures, and all this has been possible with the advent of the new technology. Even our mobiles are loaded with all similar technologies. Nevertheless, there are people who are less fortunate than us and are physically challenged, may it be deafness or being aphonic. Such people lag behind their non-handicapped peers in using these technologies. These people have some expectations from the researchers and mostly from a computer scientist that computer scientists can provide some machine/model which help them to communicate and express their feelings with others. Very few researchers have them in mind and provide their continuous works for such people.

1.2 State of Art

People normally communicate to each other by visual and sound, but the deaf and the mute can only perceive visual. This, sign language was invented to be a medium to communicate among the deaf and the mute. However, most people do not understand the sign language, creating a gap between people and the deaf for the mute, as well as difficulties in their daily-life living. [1]. Possibilities are very less in understanding the sign language of the deafdumb people by normal people in such places like

bank, booking counters etc. Conversion of sign language into text or voice reduce the problems arise there. Sign Language Recognition (SLR) is a tool executes the conversion of sign language into text or voice. Research in SLR started two decades before all over the world especially in American and Japanese Sign languages[6]. Some researchers mainly concentrated on the capture, recognition and classification of the gestures of sign language[2] English[1], Tamil[2], American[4] sign languages contributed more in this area. Many more techniques and algorithms have been proposed in using a variety of methods based on the sensor fusion signal processing, image processing, and pattern recognition methods till now[2].

II. LITERATURE REVIEW

2.2 Thai Finger-Spelling Sing Language Recognition System (TFRS)

In order to improve the quality of his/her life, research in sign language recognition has played an important role. However, sign language is not a universal language. Each country has its own variation of sign language. Even in English-Spoken countries such as the United States of America, United Kingdom, and Australia, each of them has their own creation of sign language, i.e., ASL, BSL, and Australia Sign language, respectively; the same English word may be represented by different signs. Similarly, Thai Sign Language (TSL) [1].

Sign language recognition system that receives 2-D images, 3-D images, or videos from camera generally uses image processing techniques to recognize the signs. Since cameras are inexpensive and easy to use, this type of recognition system is quite popular. However, the quality of information obtained from the camera is not good enough to accurately distinguish the characteristics of fingers, especially when fingers overlap or when occlusion occurs. This mainly leads to significant decrease in the system's performance [1].

TFRS receives data from a data glove and a motion tracker, where the data glove provides signals of finger flexures of each finger and abductions between fingers, and the motion tracker provides signals of positions and orientations. The glove is attached with 14 sensors: 10 sensors on fingers for measuring flexures and 4 sensors between fingers for measuring abductions. Figure 1 illustrates positions of sensors attached on the data glove.

However, the data glove can detect only fingers' flexures and abductions between fingers, but it cannot

provide the position and orientation of the hand. Therefore, a motion tracker is required in addition to the data glove. The motion tracker used in our TFRS is ascension 3D Guidance [1].



Figure 1. Description of sensors on Data Glove

It provides 3 values for a position in x, y, and z axes, and 3 values for an orientation on x, y, and z planes. To use the data glove and the motion tracker simultaneously, a sensor of the motion tracker is attached on then data glove [1].

TFRS receives a total of 20 data values at each timestamp, where the first 14 values are from the data glove, and the rest are from the motion tracker. The sampling rate is set to be 1,000 samples per second, with 20 days values mentioned above, per sample. Since the data from the data glove and the motion tracker is a continuous data stream, a key frame of the actual sign interval has to be identified in the data segmentation step. After the key frame is detected, it is used in the sign recognition step to identify its corresponding word [1].

2.3 Recognition of Tamil Sign Language Alphabet

In the proposed method, 32 combinations of binary images each representing ‘UP’ and ‘DOWN’ position of 5 fingers are converted into 12 vowels, 1 Aayutha Ezhuthu and 18 consonants of Tamil letters. Each image is taken through web camera in different positions for 10 times. Thus we obtained 320 images. All these images are stored in a separate directory. This may be manipulated in two types of

1. Static image Sign Recognition system.
2. Dynamic image Sign Recognition System

Feature Point Extraction

When the image is applied into Palm Extraction method, finally it changes into edge images. Every pixel in that image is ‘0’ or ‘1’. To identify the position of straight finger position of the edge image Left-Right and Right-Left scan method is used. When the scan processing started from (0, 0) position of edge image row by row up to (0,127) we get a white pixel (x₁, y₁) at any position. The difference of the height between this point and fixed reference point (x₀, y₀) colour is set in red colour at the bottom-up approach. This is calculated by using *Euclidean Distance* formula as in Euclidean distance

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

Point (x₁, y₁) is finger tip position and (x₀, y₀) is wrist hand position (Reference Point). It is considered when the height is above the particular value and stored in an array H variable and also the resized image of (x₁, y₁) is set in the colour of green. If the height is below the LEAST value, it may not be considered. Then the process continued until the capture of any white pixel up to (x₁ 1, y₁). When this is processed row by row, it may obtain 0 to 3 finger tip positions show in figure 2.

Sensor	Description
0	Thumb flexure (lower joint)
1	Thumb flexure (second joint)
2	Thumb-index finger abduction
3	Index finger flexure (at knuckle)
4	Index finger flexure(second joint)
5	Index-middle finger abduction
6	Middle finger flexure(at knuckle)
7	Middle-ring finger (second joint)
8	Middle-ring finger abduction
9	Ring finger flexure(at knuckle)
10	Ring finger flexure(second joint)
11	Ring-little finter abduction
12	Little finger flexure(at knuckle)
13	Little finger flexure (second joint)

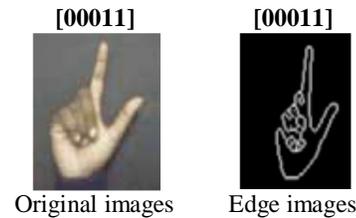


Figure 2 Sample images signs

To find out the remaining finger tip positions Right-Left scan processing can be used. After finishing scan process height values of the straight fingers in edge images are stored in an array H variable and blue colour is set in resized image of finger tip position when they are straight. To identify the straight finger (Little, Ring, Middle, Index and Thumb) from the height value, Test Phase is engaged [2].

The developed 32 binary number sign images are converted into Tamil letters which has 12 vowels, 18 consonants and 1 Aayutha Ezhuthu by image processing technique. The accuracy rate is more encouraging when compared with other regional Languages[2].

Drawbacks of existing system

The TFRS method has some disadvantage, it's used 20 sensors for data glove but any one of the sensor not working properly the sign language recognition is very difficult. The TFRS method was expensive also high because used a lot of interface. Feature Point Extraction is another method, here approaches edge detection techniques. In that case images may be detect/combine with other fingers. The accuracy of this method may be decrease.

III. PROPOSED METHODOLOGY

In this method, images are representing 'UP' and 'DOWN' position of 5 fingers are converted into 26 of English Letters. Before implement, we need Data Acquisition, Palm Image Extraction, Feature point Extraction, Training Phase and Testing Phase for the conversion. The background is to be arranged in a white colour before taking right hand palm images and the distance between the hand and web camera is to be maintained in 1 to 3 feet. The captured images are resized to a resolution of 128 X 128 pixels. It is easier to use those images for extracting the finger tip position for further processing. The way of processing showed in block diagram shows in figure 3.

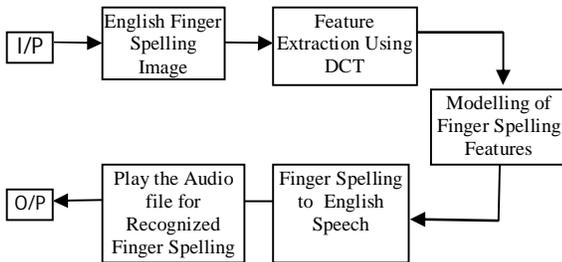


Figure 3- Block diagram for Finger spelling to English speech

a. Finger Spelling Recognition for English Alphabetic Letters

English letters also have finger spelling its way representation from letters A to Z (26 letters) shows in figure 4. For finger spelling use Index finger, middle finger, ring finger, little finger and thumb.

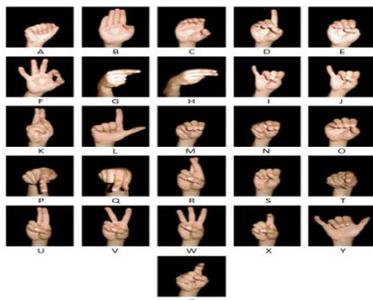


Figure 4. Screen shots for English Finger Spelling

b. Features Extraction Using DCT

Analysis of a signal is performed by using a transformation function which transform image from spatial domain to frequency domain. DCT is done by using direct application of transform to entire image or block wise. One dimensional DCT is used in audio compression method and the one dimension in audio is time. Two dimensional DCT is used in image compression, where vertical and horizontal dimensions are considered. Formulae for calculating DCT is given in equation (3) and inverse DCT is given by equation (4). DCT is used in many standardized image, audio and video compression methods. An image is subdivided into 8 X 8 block of samples. Each of these samples of the original image is mapped to the frequency domain.

Formulae of the 2-D DCT:

$$F(u, v) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} c(u)c(v)f(i, j) \left(\cos \frac{\pi(2i+1)u}{2N} * \cos \frac{\pi(2j+1)v}{2N} \right) \quad (1)$$

Formulae of the 2-D inverse DCT:

$$f(i, j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} c(u)c(v)f(u, v) \left(\cos \frac{\pi(2i+1)u}{2N} * \cos \frac{\pi(2j+1)v}{2N} \right) \quad (2)$$

where $c(u), c(v) = \begin{cases} \frac{1}{\sqrt{N}}, u, v = 0 \\ \frac{2}{N}, u, v = 1, 2 \dots N - 1 \end{cases}$

In this proposed methodology, 20 highest DCT coefficients are extracted from each finger spelling image.

b. Zigzag reordering

In this Zigzag algorithm method to scanning the images values after doing 8 X 8 DCT and quantization over a block. A new 8 X 8 blocks which denotes the value in frequency domain of the original blocks. Then it will reorder the values into one dimensional form in order to encode them. The way of representation shows in figure 5.

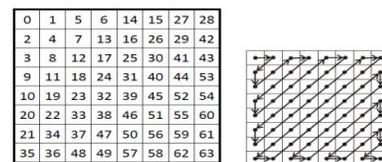


Figure 5 Zigzag representation

c. Support Vector Machines

Support vector machines (SVM) is based on the principle of Structural Risk Minimization (SRM). SVM can be used for pattern classification and nonlinear regression. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors. If the data are linearly separated, SVM trains linear machines for an optimal hyperplane and the

closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear classifier in the parameter space, but it becomes a nonlinear classifier as a result of the nonlinear mapping of the space of the input patterns into the high dimensional feature space.

d.SVM Kernel Function

SVM generally applies to linear boundaries. In the case where a linear boundary is inappropriate SVM can map the input vector into a high dimensional features space. By choosing a non-linear mapping, the SVM constructs an optimal separating hyperplane in this higher dimensional space.

Table 1. Different Types of SVM kernel functions

Types of kernels	Inner Product Kernel $K(x^T, x_i)$
Polynomial	$(x^T x_i + 1)^P$
Gaussian	$\exp\left[-\frac{\ x^r - x_i\ ^2}{2\sigma^2}\right]$
Sigmoidal	$\tanh(\beta_0 x^T x_i + \beta_1)$

IV. IMPLEMENTATION

4.1 Data acquisition

For the purpose of training the palm images, a maximum of 26 English alphabets images previously defined. For the purpose of testing, 10 sample images of some letters of the predefined signs are test at a time. For sample English letters are shown in figure 6.

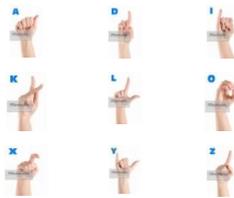


Figure 6- English letters Data Acquisition

4.2 DCT Implementation

- DCT has a strong "energy compaction" property (dimensionality reduction).
- The most of the signal information tends to be concentrated in a few low-frequency components of the DCT.
- DCT represents an image as a sum of sinusoids of varying magnitudes and frequencies

4.3 SVM Implementation

- SVM have shown superior performance than ANNs in a variety of tasks. There are two fundamental reasons
- SVM training process is guaranteed to converge to the global minimum of the associated cost function
- SVM exhibit superior generalization capability.
- This last property allows SVMs to make more accurate decision in noisy environments, which is a valuable character in the field of automatic finger spelling recognition.

4.4 Modelling of Finger Spelling Features

For training 20 highest DCT coefficients are extracted from each finger spelling and given as input to SVM model. For testing 20 highest DCT coefficient are extracted from testing finger spelling and given as input to SVM model recognize the corresponding English speech. Sample outputs English shows Input, SVM trained files and output in audio files in figure 7.

Sample Output for English letters

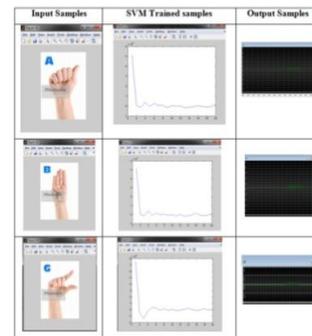


Figure 7. Sample output for finger spelling

V. EXPERIMENTAL RESULTS

The proposed system is implemented using a MATLAB program where it is evaluated for finger spelling recognition.

5.1 Finger Spelling image converted into to English Speech

- For Training 26 Finger Spelling English images and 20 highest DCT coefficients are extracted from each finger spelling and given as input to SVM model. SVM is used recognize the sign and display corresponding English speech.
- For Testing 10 Finger Spelling English images and 20 highest DCT coefficients are extracted from each finger spelling and given as input to SVM model. SVM is used recognize the sign and display corresponding English speech.

- Performance of finger spelling to English speech using DCT and SVM is given in the table 3 and 4.
- In a SVM training to change the kernel functions polynomial, Gaussian, Sigmoidal. The table shows the various of recognition Rate for each kernel functions.
- The bar chart shows comparison of SVM kernel functions for finger spelling English speech refer figure 8.
- From the bar chart (figure 8) shows highest recognition rate 94.11% for Gaussian.

Table 3. Comparison SVM kernel functions for English

Sl.No	SVM Kernel	Recognition Rate
1.	Polynomial	90.32%
2.	Gaussian	94.11%
3.	Sigmoidal	93.28%

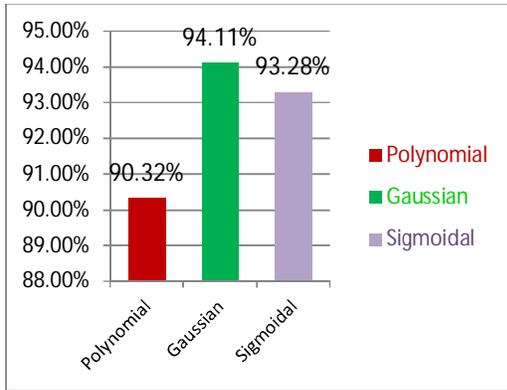


Figure 8- SVM kernel functions representation in bar chart

VI. PERFORMANCE ANALYSIS

6.1 Training Phase for English Alphabet letters

This phase is designed to improve the recognition rate of the alphabet gestures, (A to Z), in previously done works. Nine alphabets are chosen for to show the performance evaluation. These are alphabet A, alphabet D, alphabet I, alphabet K, alphabet L, alphabet O, alphabet X, alphabet Y.

In table 4 contains Alphabet letters, Gesture and corresponding recognized gesture, recognized rate and time.

Table 4. Training phase for English letters

Sl.No	Alphabet	Gesture	Recognized Rate	Recognized Time (seconds)
1.	A		100%	0.057
2.	D		100%	0.034
3.	I		100%	0.067
4.	K		100%	0.047
5.	L		100%	0.040
6.	O		100%	0.033
7.	X		100%	0.037
8.	Y		100%	3.347
9.	Z		100%	2.297

In training phase all the English alphabet letters are recognized successfully. But showed some sample letters only in above table.

6.2 Testing Phase for English Alphabet letters

For testing phase, five alphabets are chosen. These are alphabet A, alphabet B, alphabet I, alphabet L, alphabet O. In table 5 contains Alphabet letters, Gesture and corresponding recognized gesture, recognized rate and time using number of samples.

Table 5. Testing phase for English letters

Sl.No	Alphabet	Gesture	No. of samples for testing	No. of false result patterns	Accuracy in Percentage	Recognized Time (seconds)
1.	A		10	0	100	0.057
2.	B		10	0	100	0.934
3.	I		10	0	100	0.067

4.	L		10	1	90	0.040
5.	O		10	2	80	0.033

- Stephan Liwicki and Mark Everingham, 'Automatic Recognition of Fingerspelled Words in British Sign Language', *IEEE workshop*, pp.1-9, 2009.

The overall accuracy of the proposed method for English finger recognition.

$$\text{Accuracy} = \frac{\text{No.of patterns} - \text{No.of false patterns}}{\text{No.of patterns}} \times 100$$

The overall performance of the training phase is given accuracy 100% and testing phase is given accuracy 94%.

VII. CONCLUSION

An English finger spelling recognition system proposed for human computer interaction using Image processing Technique and SVM Neural networks was implemented successfully with accuracy also comparable. It was recognized 26 English Alphabetic letters and identified the finger spelling images which was presented in the English speech. 20 highest DCT coefficients were extracted from each finger spelling and given as input to SVM model. SVM is used to recognize the sign (finger spelling) and display corresponding English speech. A part of the specified SVM soft computing techniques the work can be extended to add some other softcomputing techniques and also to increase the performance and accuracy of the Sign language better than this approach.

VIII. REFERENCES

- Supawadee Saengsri, Vit Niennattrakul, and Chotirat Ann Ratanamahatana, 'TFRS: Thai Finger-Spelling Sing Language Recognition System', *IEEE*, pp: 457- 462, 2012.
- Subha Rajam, P. and Balakrishnan, G., 'Recognition of Tamil Sing Language Alphabet using Image Processing to aid Deaf-Dumb People', *International Conference on Communication Technology and System Design*, Elsevier, pp: 861-868,2011.
- Subha Rajam, P. and Balakrishnan, G., 'Real Time Sign Language Recognition System to aid Deaf-dumb People', *IEEE*, pp: 737-742,2011.
- Chance, M. Glenn, Divya Mandloi, Kanthi Sarella, and Muhammed Lonon, (2005) 'An Image Processing Technique for the Translation of ASL Finger-Spelling to Digital Audi or text', *NTID International Instructional Technology and Education of the Deaf Symposium*, pp: 1-7.
- Nicolas Pugeault, and Richard Bowden, 'Spelling It Out: Real-Time ASL Fingerspelling Recognition', *IEEE Workshop on Consumer Depth Cameras for Computer Vision*, pp. 1-6, 2011.
- Deepika Tewari, and Sanjay Kumar Srivastava, 'A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm' *IJEAT*, pp. 165-170,2012.
- Sakshi Goyal, Ishita Sharma, and Shanu Sharma, ' Sign Language Recognition System for Deaf and Dumb People', *International Journal of Engineering Research & Technology*, pp. 382 – 387, 2013.
- Pingale Prerna Rambhau, 'Recognition of Two Hand Gestures of word in British Sign Language (BSL)', *International Journal of Scientific and Research publications*, pp. 1-5, 2013.