ISSN: 1748-0345 (Online) www.tagajournal.com

Real Time Hand Gesture Recognition System

Neethu P S[1, 1*]
[1]Research Scholar, Dept. of Information & Communication Engineering,
Anna University, Chennai
[1*]Assistant Professor, Dept. of ECE, New Prince Shri Bhavani College of Engineering & Technology,
Chennai

Dr. R Suguna[2]
[2]Professor, Dept. of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai

Dr. Divya Sathish[3]
[3]Professor, Dept. of CSE, SKR Engineering College,
Chennai

***Abstract*-- In this paper, automatic detection and classifications of hand sign language recognition system is proposed. It consists of preprocessing, transformation, feature extraction and classifications. The RGB image is converted into grey scale image as a preprocessing method and then this image is converted into multi resolution image using Gabor transform. Then, features are extracted from Gabor transformed image and these features are classified using Adaptive Neuro Fuzzy Inference System (ANFIS) classifier. The performance of the proposed hand sign language detection system is analyzed interms of sensitivity, specificity, accuracy, and classification rate and detection time.**

***Keywords: Hand sign, feature extraction, classification, Gabor, multi resolution*.**

I. INTRODUCTION

Nowadays, Pattern Recognition (PR) is an emerging filed in and around the world to distinguish the various patterns from one region to another region. Sign language is one kind of this PR technique which represents the thinking of human beings in nature. It is very much useful for deaf and dump disable persons to recognize the language of the normal people around the world. The sign language can be generated using hands, limbs, lips and face. It is one kind of communication which has no sound format to deaf and dump persons. Around the world, the common well known sign languages are American Sign Language (AMSL), British Sign Language (BSL) and Australian Sign Language (ASL). Most of the people are using BSL language pattern to communicate with disables persons.

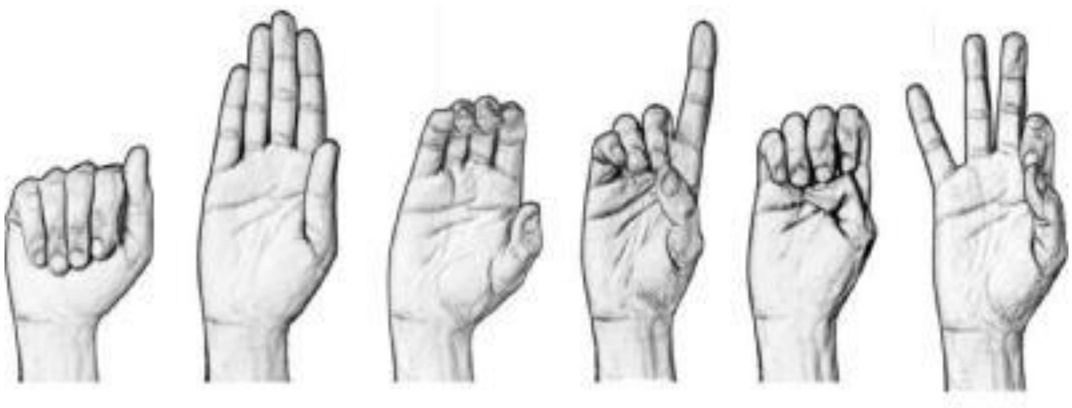


Figure 1: Examples of sign languages generated by hands

Fig.1 shows the example images of various sign languages which are generated by hands In past decades, sign languages are created using hands with the help of kinetic sensor which is placed over the head. The main limitation of this method is its cost and it is not suitable for all set of peoples. Another limitation is that tha size and shape of various hand signs are similar in nature which makes the classification process more complex as illustrated in Fig.2.

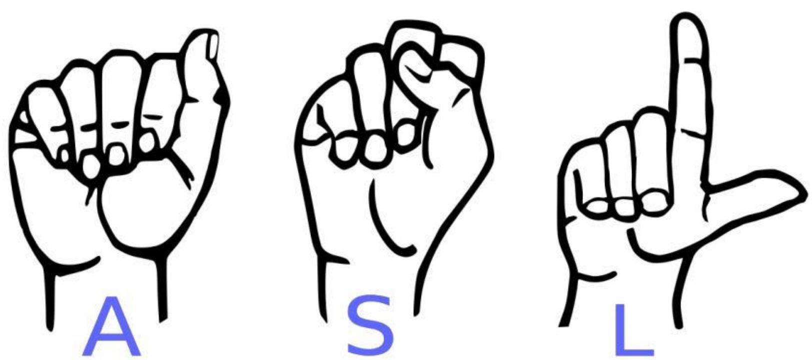


Figure 2: Similar hand sign language images

In Fig.2, the first image and second image is having very similar pattern which makes detection and classification process more complex. Hence this paper proposes a methodology to detect and classify the hand generated sign languages using image processing techniques without kinetic and other sensors.

II. LITERATURE SURVEY

(Funde et al, 2016) used Row mean transform and Cosine-Haar Hybrid Wavelet Transform generated using assorted proportion of constituent Transform for the detection of hand sign languages. The authors deliberated about slope magnitude method for detecting edge and multiple transform for energy composition and also extracted the feature by using row mean of column transformed image, going ahead with the help of these edge image and various transform calculate the genuine acceptance ratio. (J. Wu et al, 2016) (Nikam et al, 2016) used Support Vector Machine (SVM) classification approach for detecting and classifying various sign languages in real time mode under different environmental conditions. The authors achieved 89% of classification rate as an average value for hand sign languages. (Nachamai et al, 2013) used Scale Invariant Feature Transform (SIFT) technique for identifying the gestures of the deaf and dump persons. This transformed produced images which had both scale and orientations. This proposed method was tested on different set of sign language dataset under various illumination conditions.

(Jayashree et al, 2012) proposed a methodology for hand gesture recognition using Euclidian distance metric algorithm. The authors classified both static and dynamic hand gesture images in an affective manner with respect to time complexity and classification accuracy. The authors achieved 90.1% of recognition rate using distance metric technique.

III. PROPOSED METHODOLOGY

In this paper, automatic detection and classifications of hand sign language recognition system is proposed. It consists of preprocessing, transformation, feature extraction and classifications. The proposed flow of hand sign language recognition is illustrated in Fig.3.

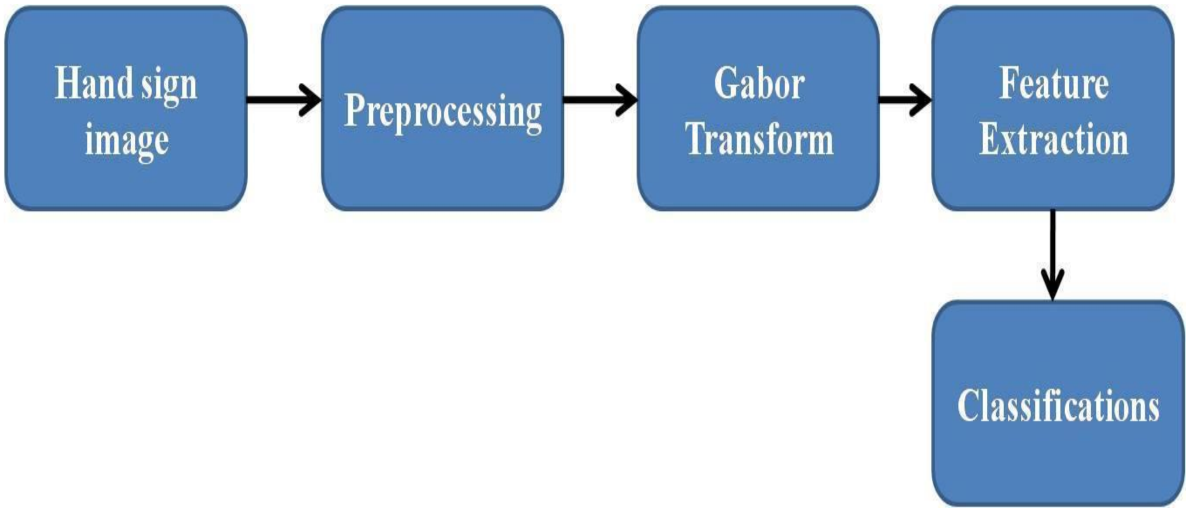


Figure 3: Proposed hand sign language recognition system

A. Preprocessing

It is essential step for improving the classification rate. In this paper, all sign language images which are obtained from open access dataset are RGB color format. Hence, each pixel in this image has 24 pixels and its processing time is high. Hence, this RGB image is converted into grey scale image in which each pixel has only 8 bits in nature, which increases the computational time period. Then, all the grey scale converted images are resized into 128 *128 pixels as rows and columns.

B. Multi resolution transform

Transformation is the process of converting the spatial domain image into frequency domain image. This is process is categorized into single resolution and multi resolution transformation. In case of single resolution transformation, spatial domain image is converted into frequency domain, where as, spatial domain image is converted into multi resolution (spatial, frequency and amplitude components) image in case of multi resolution transformation. Hence, the accuracy level is high by using multi resolution transformation. Contourlet and curvelet transforms were used as conventional multi resolution transforms. The conversion process is complex in these transforms due to the large number of sub bands. Hence, Gabor transforms is used in this paper which performs multi resolution process with out using sub bands.

The preprocessed sign language image is convolved with Gabor kernel inorder to produce the Gabor transformed image. In this work, 2D Gabor kernel with different scale and orientation is used as stated in the following equation.

$$G(x, y, \theta, f) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} \times \exp\{2\pi\ (f * xb * \cos\theta + f * yb * \sin\theta)\} \qquad (1)$$

Where as, Gabor kernel is denoted by G and it is based on the terms orientation (θ) and scale (f). The pixels in preprocessed sign language image is represented by (x, y) and its variance is given by σ. The parameters xb and yb in Equation (1) are given as,

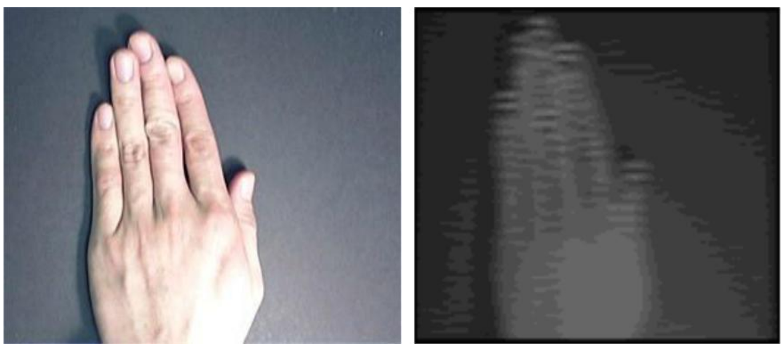
$$xb = x^* \cos\theta + y^* \sin\theta \qquad (2)$$

$$yb = x^* \cos\theta - y^* \sin\theta \qquad (3)$$

In this paper, the scale value are varies as f= {1, 2, 3, 4, 5} and the orientation varies from 0^0 to 180^0 by step of 20^0. Hence, Gabor kernel is produced for every scale with respect to eight orientations. Finally, 40 Gabor kernels are produced by varying the scale value. Now, the preprocessed sign language image is convolved individually with these 40 Gabor kernels which produces 40 Gabor transformed images. Then, the magnitude of each pixel in 40 Gabor transformed image are computed using the following equation as,

$$|G| = \text{sqrt}\ (\text{Real}\ (G)^2 + \text{Imaginary}\ (G)^2) \qquad (4)$$

Now, we have 40 Magnitude Gabor transformed images which has only real terms. Then, the maximum of each pixels in Magnitude of Gabor transformed images is computed which produces single Gabor transformed magnitude image. This image contains amplitude, frequency and orientation. In this paper, orientation image is not used due to its overlapping of orientation levels in the preprocessed image.



(a) **(b)**

Figure 4: (a) Test hand sign language image (b) Gabor magnitude image

Fig. 4 (a) shows the test hand sign language image and Fig.3 (b) shows the Gabor magnitude image.

C. Feature Extraction

Features are the essential components of the objects in image. They differentiate the various regions within the image. In this paper, Discrete Wavelet Transform (DWT), Local Binary Pattern (LBP) and Grey Level Coocuurence Matrix (GLCM) features are extracted from the Gabor transformed magnitude image of the sign language image. Hence, total 4 sub bands from DWT, one feature from LBP and four features from GLCM which accumulates overall nine features used in this paper.

D. DWT features

In this paper, two dimensional DWT transformed is applied on the Gabor magnitude image, which transforms the source image into four number of sub bands. Each sub band reflects the characteristics of the image with respect to its different orientations. The sub band formation in DWT transform is explained in Fig.5. The Gabor transformed sign language image is applied to Low Pass Filter (LPF) and High Pass Filter (HPF), respectively. Each filtered image is now decimated by sampling factor 2 to overcome the sampling errors. These decimated images are again passed through LPF and HFP filters with decimator which produces four sub bands as shown in Fig.5.

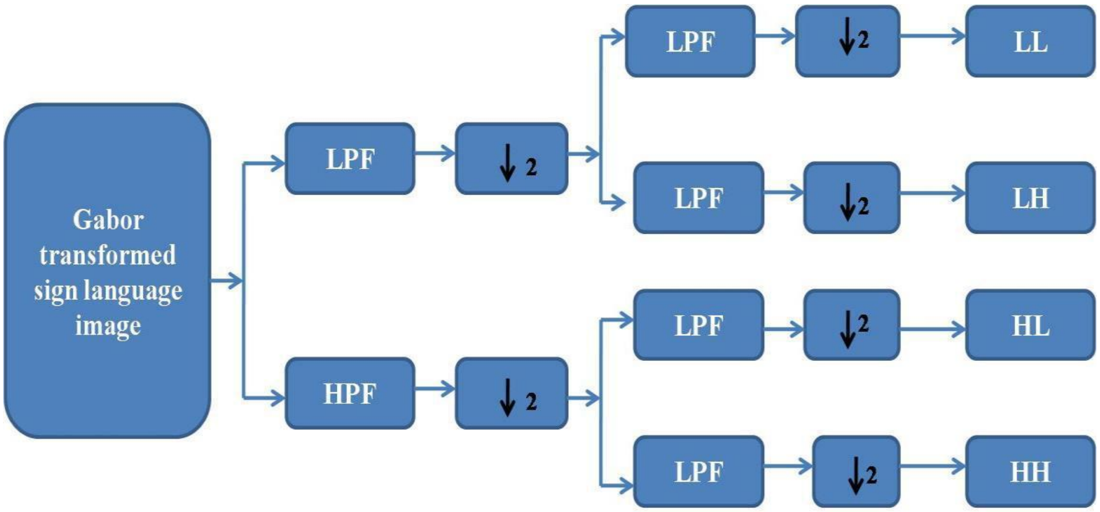
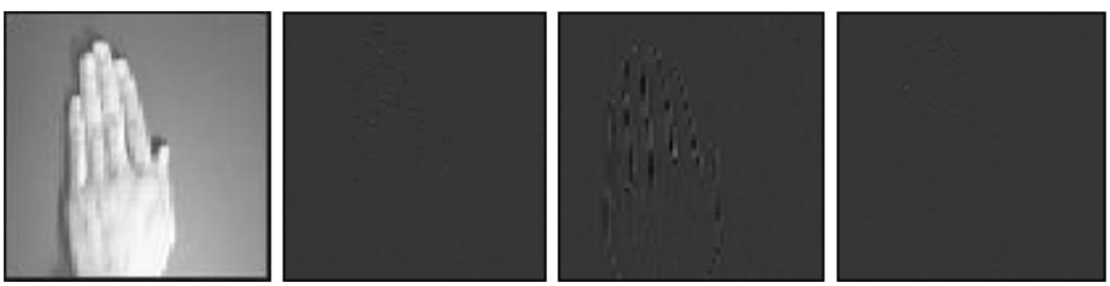


Figure 5: Formation of sub bands using DWT

These four sub bands are 'Approximate' which is illustrated as LL, second sub band is ' Horizontal' which is represented as 'LH', third sub band is 'Vertical' which is represented as 'HL' and fourth sub band is 'Diagonal' which is represented as 'HH'. These four sub bands are depicted in Fig.6 (a)-Fig.6 (d). LL sub band has low frequency factors and HH sub band has high frequency factors.



(a) (b) (c) (d)

Figure 6: (a) Approximate band (b) Horizontal band (c) Vertical band (d) Diagonal band

In this paper, the size of source image is 128*128 pixels as width and height. DWT transfoms the source image into four equal size sub bands. Hence, size of DWT extracted feature is 4*128*128 pixels.

E. LBP

The binary characteristics of the sign language image are defined by LBP features, which is single metric feature derived from single image. The size of LBP feature image is equals to the size of the source image, where LBP features are to be derived. In this paper, the size of Gabor transformed magnitude image is 128*128 as width and height format. Hence, the size of LBP feature is also 128*128 as width (P) and height (R). Each pixel in Gabor magnitude image produces LBP feature metric, based on its eight surrounding pixel values. The functional representation of LBP feature extraction over Gabor transformed image is given in the following equation as,

$$\mathrm{LBP}_{P,R} = \sum_{p=1}^{P} 2^{(p-1)} \times f_1(g_p - g_c)$$
$$f_1(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{else} \end{cases} \tag{5}$$

In this equation, g_p if surrounding pixels and g_c is the center pixel where the feature metric to be extracted. The extracted LBP feature image is shown in Fig.7.

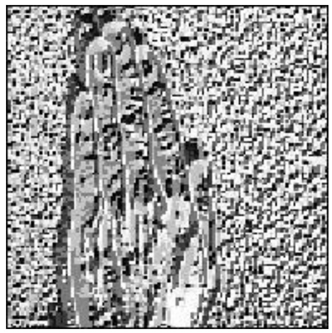


Figure: 7 LBP sign language image

In this paper, the size of extracted LBP feature is 128*128 pixels as width and height.

F. GLCM features

The texture features are extracted from Gabor transformed image for improving the sign language classification rate with respect to its intensity levels. In this paper, GLCM features are extracted from Gabor magnitude sign language image as texture features which exhibit the intensity variations in the image. The GLCM features are derived from GLCM matrix, which can be constructed at the orientation of 45^0. The texture features as Contrast (C), Energy (E), Homogeneity (H) and Correlation (CORR) are derived from GLCM matrix as depicted in the following equations.

$$C = \sum(|r - s|^2 \times G(r, s) \tag{6}$$

$$E = \sum G(r, s)^2 \tag{7}$$

$$H = \frac{\Sigma(,)}{1+|-|} \tag{8}$$

$$CORR = \sum(r - \mu)(s - \mu) \frac{G(r,s)}{[\sigma r . \sigma s]} \tag{9}$$

Where as, the number of rows and coloums in GLCM matrix (G) is represented by 'r' and 's' and σr and σs are the standard deviations of the GLCM matrix with respect to its row and coloumn, respectively.

The size of GLCM feature is 4 elements in this paper. All the extracted features in this paper are assembled as matrix format and feed to the classifier for improving the classification rate.

G. Classifications

The main objective of this paper is to detect the sign language image in an effective manner. In this regard, classifier plays an important role in this detection process. Many conventional classifiers such as Neural Networks (NN), Support Vector Machine (SVM) and random forest classifiers were used by many researchers for the sign language detection process. The classification rate was low due to its complexity nature with respect to improper training and testing samples. To eliminate such drawbacks, ANFIS classifier is adopted in this paper to improve the sign language classification rate with proper training and testing samples. The ANFIS classifier used in this paper have 5 internal layers for training and testing the samples. Each internal layer is having fixed number of nodes as fixed as 10 after several testing to obtain maximum efficiency. The internal architecture of the ANFIS classifier is shown in Fig.8. The same architecture can be used for both training and testing of the various sign language images.

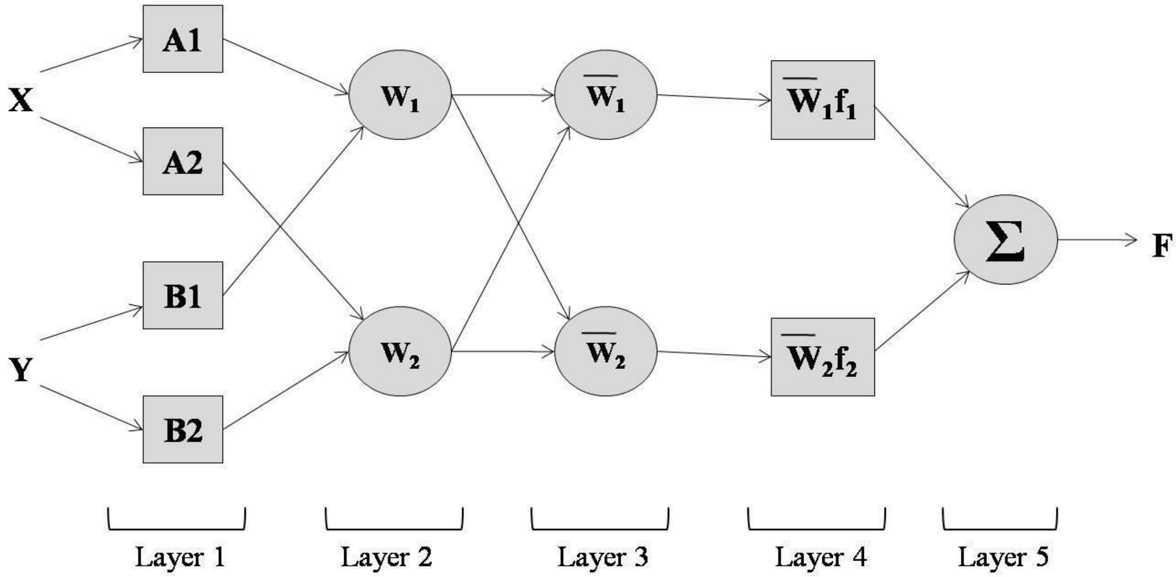


Figure 8: Training and Testing of ANFIS classifier

The weights of the layers in ANFIS classifier are represented as W1 and W2, respectively. These weights can be computed using the following equation as,

$$W_i = \text{Mean}(X) * \text{Mean}(Y) \qquad (10)$$

During training mode of the ANFIS classifier, X and Y represent the features training samples from the well known sign language images. During testing mode of the ANFIS classifier, X represents the trained patterns and Y represents the features which are extracted from the source sign language image. The inverse of the weighting functions (W1 and W2) are determined as,

$$(\ 1) = \ 1 = \frac{1}{1+2} \qquad (11)$$

$$(\ 2) = \ 2 = \frac{2}{1+2} \qquad (12)$$

The final output of the ANFIS classifier is represented using the following equation as,

$$F = \frac{\Sigma \ 1*1}{\Sigma} \qquad (13)$$

The output pattern from this tested ANFIS classifier is flipped between the values from 0 and 1. If the number of tested sign language images is 10, then the output value of this ANFIS classifiers are varies from 0 to 1, which stepped at 0.2. Fig.9 shows various test sign language images used in this paper.

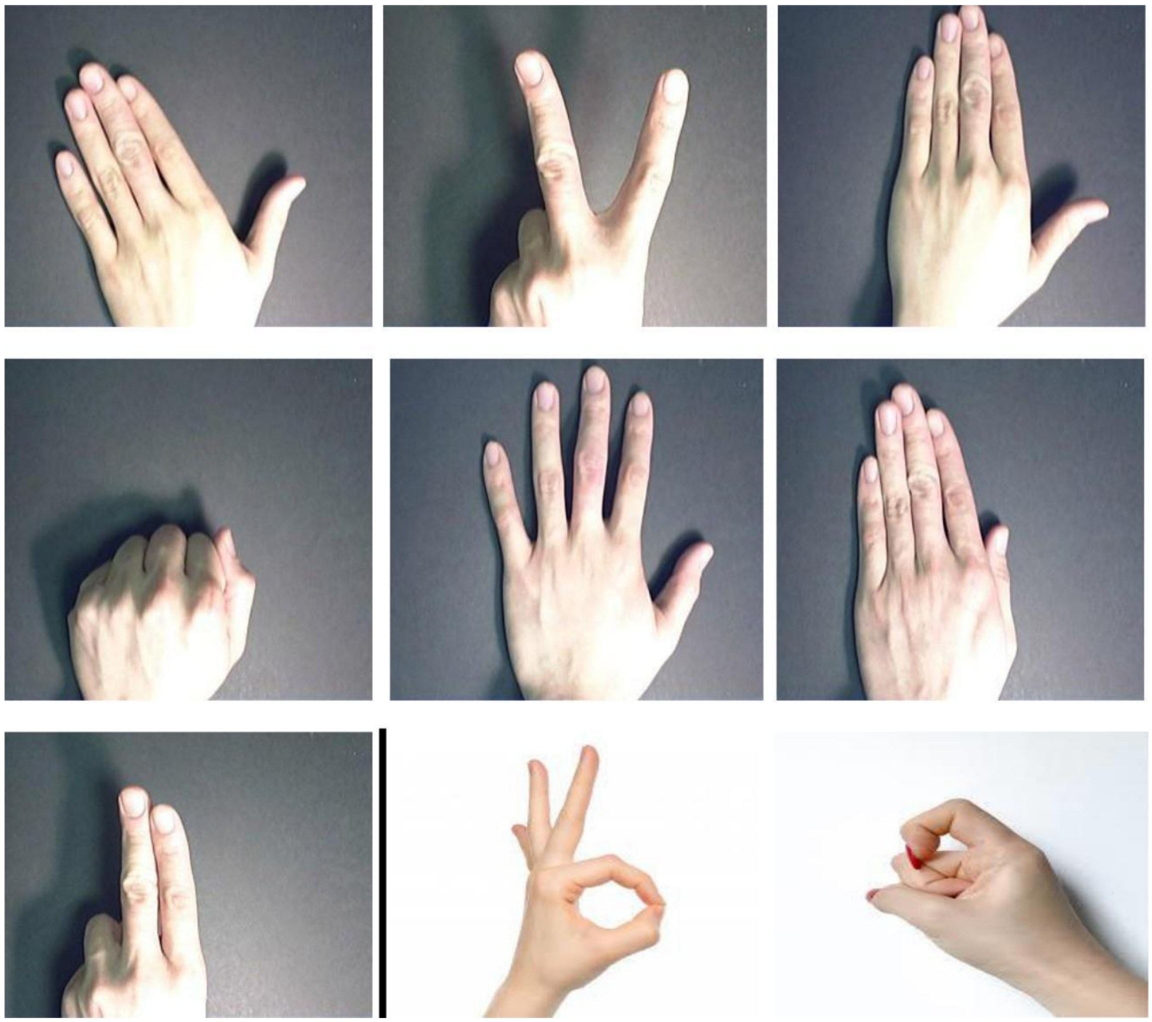


Figure 9: Test hand sign language images

IV. RESULTS AND DISCUSSION

In this paper, MATLAB R2014b is used to simulate the proposed sign language classification system, with 2GB RAM and Intel Pentium core 2 duo processor as hardware specifications. The proposed methodology for sign language classification is applied on the sign language images which are obtained from American Sign Language Lexicon Video Dataset (ASLLVD) dataset. This dataset is open and no licence is required to use the images from this dataset. This open access dataset contains 2,284 monomorphemic lexical sign images and each image has the size of 128 and 128 pixels as width and height. The performance of the proposed hand sign language detection system is analyzed interms of sensitivity, specificity, accuracy, and classification rate and detection time.

( ) = /(+) (14)

( ) = /(+) (15)

( ) = (+)/(+ + +) (16)

= / (17)

Where as, TP is Trus Positive which indicates the total number of correctly detected sign language images and TN is True Negative which indicates the total number of correctly detected non-sign language images. FP is False Positive which indicates the total number of wrongly detected sign language images and FN is False Negative which indicates the total number of wrongly detected non-sign language images. Table 1 shows the performance analysis of proposed sign language detection system interms of sensitivity, specificity and accuracy. The proposed methodology stated in this paper achives 92% of sensitivity, 97% of specificity, 98% of accuracy and 97% of classification rate.

Table 1 Performance analysis of proposed methodology

Methodology	**Experimental Results (%)**
Sensitivity	92
Specificity	97
Accuracy	98
Classification rate	97

Fig.10 shows the graphical plot of performance evaluation parameters of proposed method.

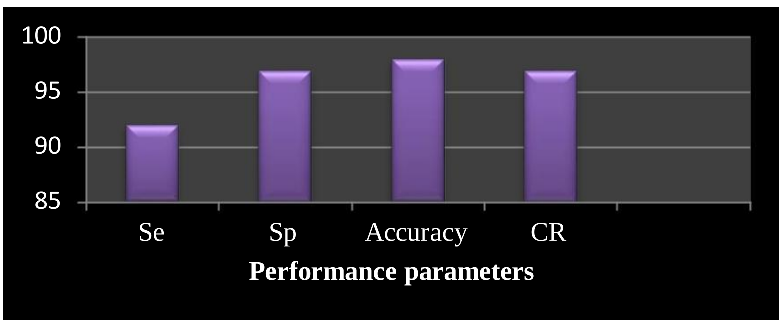


Figure 10: Graphical plot of performance evaluation parameters of proposed method

Table 2 Impact of features on hand sign language classifications

Features Index	**Classification rate (%)**
1	89
1 and 2	92
1,2 and 3	97

*feature index of LBP is 1; feature index of GLCM is 2; feature index of DWT is 3;

The proposed hand gesture detection and classification system achieves 89% of classification rate while using LBP features alone. The classification rate of the proposed hand gesture detection system is 92% while the feature index1 integrated with feature index2. The classification rate of the proposed hand gesture detection system is 97% while the feature index1 integrated with feature index2 and feature index3, as shown in Table 2.

Table 3 shows the performance comparisons of proposed methodology with state of arts interms of sensitivity, specificity and accuracy.

Table 3 Performance comparisons of proposed methodology with state of art

Methodology	Sensitivity	Specificity	Accuracy
Proposed Methodology	92	97	98
Yang et al. (2017)	85	92	93
Wu et al. (2016)	87	94	91
Quan and Liang (2016)	86	93	92

The proposed hand sign language recognition system is compared with conventional hand gesture recognition systems as Yang et al. (2017), Wu et al. (2016) and Quan and Liang (2016). The conventional methodology Yang et al. (2017) achieved 85% of sensitivity, 92% of specificity and 93% of accuracy; where as Wu et al. (2016) achieved 87% of sensitivity, 94% of specificity and 91% of accuracy and Quan and Liang (2016) achieved 86% of sensitivity, 93% of specificity and 92% of accuracy, as depicted in Table 3.

V. CONCLUSION

In this paper, automatic detection and classifications of hand sign language recognition system is proposed. The source sign language hand image which is in the form of spatial mode is converted into multi resolution image using Gabor transform. Then, features are extracted from Gabor transformed image and these features are classified using ANFIS classifier. The performance analysis of proposed sign language detection system interms of sensitivity, specificity and accuracy. The proposed methodology stated in this paper achives 92% of sensitivity, 97% of specificity, 98% of accuracy and 97% of classification rate.

References

[1] http://www.bu.edu/av/asllrp/dai-asllvd.html

[2] (Ghotkar, A. S. et al, 2012) *Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction*, Int. J. of Human Computer Interaction, **vol 3** (no.1) pp.15 – 25.

[3] (Gurjal, P et al, 2012) *Real time Hand Gesture Recognition using SIFT*, Int. J. of Electronics & Electrical Engg. pp. 19 – 33.

[4] (Jayashree R. Pansare et al, 2012) *Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background*, Journal of Signal and Information Processing, pp. 364-367.

[5] (Kishore, P. V. V. et al , 2012) *Segment, Track, Extract, Recognize and Convert Sign Language Videos to Voice/ Text*, Int. J. of Advance Comp. Sc. and Applications, pp.35 – 47.

[6] (Kristensson, P. O. et al, 2012) *Continuous Recognition of One-handed and Twohanded Gestures using 3-D Full-body motion tracking sensors,* In Proc. of IUI 12.

[7] (Nachamai. M et al, 2013) *Alphabet Recognition of American Sign Language: A hand gesture recognition APPROACH using SIFT algorithm*, International Journal of Artificial Intelligence & Applications (IJAIA), **Vol.4**, (No.1).

[8] (Nikam A S. et al,2016) *Sign language recognition using image based hand gesture recognition techniques*, Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, pp. 1-5.

[9] (Quan C et al, 2016) *A Simple and Effective Method for Hand Gesture Recognition,* International Conference on Network and Information Systems for Computers (ICNISC), pp. 302-305.

[10] (Rautaray et al, 2012) *Real time Hand Gesture Recognition System for Dynamic Applications,* Int. J. of UbiCom, pp.21 – 31.

[11] (Rautaray, S. S. et al, 2012) *Real time Multiple Hand Gesture Recognition System for Human Computer Interaction,* Int. J. of Intelligent Systems and Applications, pp. 56 – 64.

[12] (Sultana, A. et al, 2012) *Vision based Gesture Recognition for Alphabetical Hand Gestures using the SVM Classifier*, Int. J. of Comp. Sc. And Engg. Tech. **vol 3** (no.7) pp.218 – 223.

[13] (Wu. J et al, 2016) *A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors*, IEEE Journal of Biomedical and Health Informatics, vol. 20,(no. 5), pp. 1281-1290.

[14] (Yang J et al, 2017) *Real time hand gesture recognition via finger-emphasized multi-scale description*, IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, Hong Kong, pp. 631-636.

[15] (Funde A.G et al, 2016) *Improved sign language recognition using Cosine-Haar hybrid wavelet transform and assorted similarity measures with energy compaction, International Conference & Workshop on Electronics & Telecommunication Engineering (ICWET 2016)*, pp. 174-177.