Volume 5, Issue 5, May 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Automated System for Indian Sign Language Recognition

Chandandeep Kaur, Nivit Gill

Department of Computer Science,
Punjabi University,
Punjab, India

Abstract*— Sign Language is a non-verbal communication mode which is used by the people with hearing disabilities, who cannot use voice sounds for communication. Instead, they need signs or actions to express themselves. Normal people are usually unaware of these signs. Therefore, the need of a computer based intelligent system is highly demanding for the dumb community that will enable them significantly to communicate with all other people using their natural hand gestures. This paper presents a method for automatic recognition of signs on the basis of shape based features. For segmentation of hand region from the images, Otsu's thresholding algorithm is used, that chooses an optimal threshold to minimize the within-class variance of thresholded black and white pixels. Features of segmented hand region are calculated using Hu's invariant moments that are fed to Artificial Neural Network for classification. Performance of the system is evaluated on the basis of Accuracy, Sensitivity and Specificity.

***Keywords*— Indian Sign Language (ISL), Artificial Neural Network (ANN), Gesture Recognition, Hu's invariant moments, Histogram Equalization**

I. INTRODUCTION

Language is a human system of communication that uses arbitrary signals, such as voice sounds, words as well as gestures to express inner thoughts and emotions. Accordingly, it can be classified as: Verbal and Non-verbal [1]. The act of transferring messages between individuals using speech is known as verbal communication. While, the act of expressing your ideas with the help of facial expressions or gestures is as non-verbal communication.

In the definition of language, it is necessary to be careful not to exclude symbols, gestures, or motions as it is the language of deaf community who cannot use voice or sound for communication. Instead, they need sign language. It is a non-verbal communication mode that is the combination of orientation and movements of hands, arms or body, and facial expressions [2]. There are various sign languages used all over the world. In India, deaf community uses Indian Sign Language (ISL), which has its own with specific syntax, grammar, morphology and phonology. In India, there is a huge variation in its culture, language and religion [3], due to which standard form of ISL has not developed yet.

In January 1999, the Ramakrishna Mission made an effort to standardize ISL by creating a sign dictionary. The first 'Indian Sign Language Dictionary' was released in August 1999. It provides a common sign language code for about 2500 signs from 42 cities of 12 states [3].

Sign Language consists of two major components: Finger spelling, provides a unique sign for each letter of alphabet (A-Z) and numerals (0-9). And, word-level gestures, provides corresponding sign for each word of vocabulary [4].

The problem arises when deaf or dumb people try to express themselves to normal people with the help of these sign language grammars. But normal people are usually unaware of these signs and very few of these are recognized by them. As a result, communication of dumb person is limited within his/her family only [5].

Therefore, at this age of technology, the demand for a computer based intelligent system is highly demanding for the dumb community that will enable them significantly to communicate with all other people using their natural hand gestures. It will provide opportunities for them in Industry Jobs, IT sector Jobs, and Government Jobs.

However, a lot of work has been done on interpretation of sign languages; a little work was done on Indian Sign Language. Also, the initial work was carried out over small dataset which acquire large amount of data processing time [6]. So, there is a need to design new algorithm for recognizing a set of commonly used hand gestures in ISL.

To recognize sign language, two approaches can be followed, Sensor based and Vision based. In Sensor based approach, a sensing device needs to be attached to user that measures the joint angles, position of fingers and hands [7][8]. Whereas, vision based approach works directly on image gestures captured by camera [9]. Figure 1 shows example of sensor based and vision based approaches.

The rest of paper is organized as follows: Section II discusses about the previous research work done in the same area. In section III, proposed methodology is explained. Section IV discusses experimental results obtained and finally, Section V concludes the paper along with the future work.

Figure1(a): Sensor based approach

Figure1(b) Vision based approach

II. Literature Review

A lot of work has been done on developing systems for different sign languages. **D. K. Ghosh and S. Ari** proposed a vision based approach that uses histogram based thresholding algorithm for segmentation of hand region [8]. A Localized Contour Sequence (LCS) feature is considered for classification using k-mean based radial basis function neural network. **P. R. Futane and R. V. Dharaskar** [3] uses Fourier Descriptors to extract shape and geometry features. System is trained by general purpose Fuzzy MinMax neural network that gives 92.92% accuracy. An automatic ISL recognition system is proposed by **K. Dixit and A. S. Jalal** [7] that employs the use of Hu invariant moments in combination with structural shape descriptors to form new feature vectors. Multi class Support Vector Machines (MSVM) is used for classification. **K. Dabre and S. Dholay** worked on a machine learning model that uses image processing and neural network methodologies to characterize hand images taken from video through camera [10]. **N. Baranwal and N. Singh** introduced a Novel hand gesture recognition technique using Discrete Wavelet Packet Transform (DWPT) [6]. The Principal Component Analysis (PCA) is used for extracting the significant features for classification using artificial neural network. **A. Nandy and S. Mondal** [11] uses Hidden Markov Model (HMM) and Bhattacharya Distance estimation for classification that uses features from orientation histogram. A sensor based system for American Sign Language is proposed by **K. F. Li and K. Lothrop.** X-Box Kinect is used as a gesture input device that captures 3D data of joints [2] which is analysed and compared with pre-recorded signs using 3D point pattern matching algorithm. Work of **S. Atif and Y. N. Khan** explores the use of sensor gloves in Sign language recognition [12]. ANN recognize the sensor values coming from the sensor glove and categorize them into 24 alphabets and two punctuation symbols with 88% accuracy. **C. Oz. and M. C. Leu** [13] uses flock of Birds 3-D motion tracker to extract features that are used by multilayer ANN to recognize signs.

III. Proposed Methodology

In this paper, a method to classify various static and dynamic hand gestures used in ISL is proposed. Figure 2 shows the steps of proposed method.

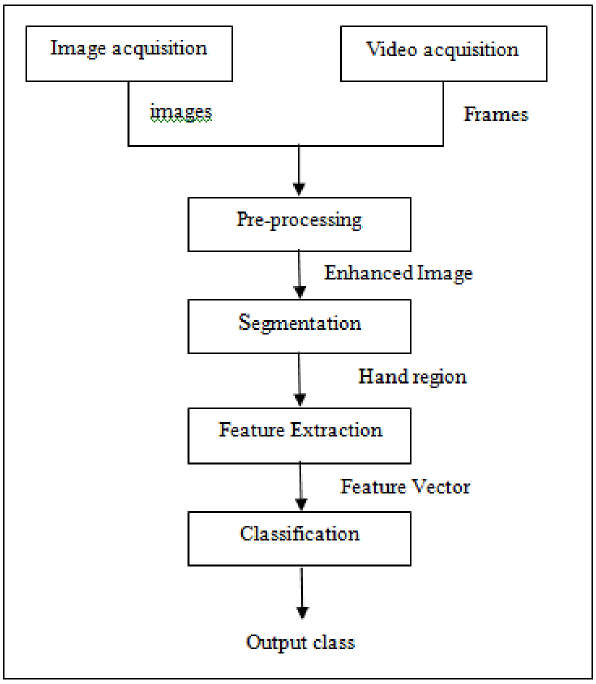

Figure 2: Flowchart showing steps of Sign language recognition

A. Pre-processing

As the dataset acquired was created under different lightning conditions, the images have different contrast values. So, there is a need to adjust the pixel intensity values before further processing. The low-contrast image's histogram is narrow and centered toward the middle of the gray scale. If we distribute the histogram to a wider range the quality of the image will be improved, as the histogram of high contrast image covers broad range of the gray scale and the distribution of pixels is not too far from uniform, with very few vertical lines being much higher than the others. For this purpose, histogram equalization is applied that adjust and normalize brightness and contrast of processing images [15]. The result of pre-processing is shown in Figure 3.

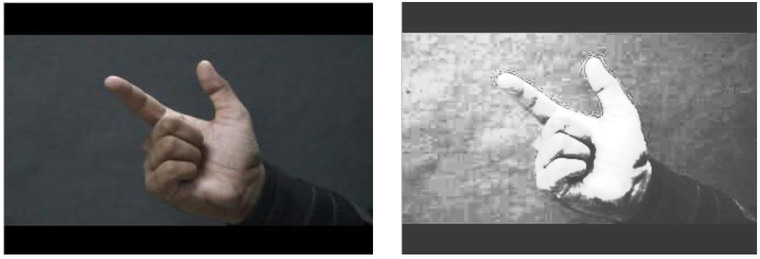

Figure 3: Results of Pre-processing

B. Segmentation

Image segmentation is the process of partitioning an image into meaningful and non-overlapping regions or objects [16]. Region of interest is extracted from the whole image for further analysis. Among various segmentation approaches, Otsu's thresholding method is used in the proposed method to extract hand region from background, as it is known to be one of the most successful method for image segmentation. It has advantage of simple and fast calculations. It is region based segmentation method with automatic threshold selection. It selects an optimal gray level threshold value, based on which hand region is extracted. Segmentation results are shown in Figure 4.

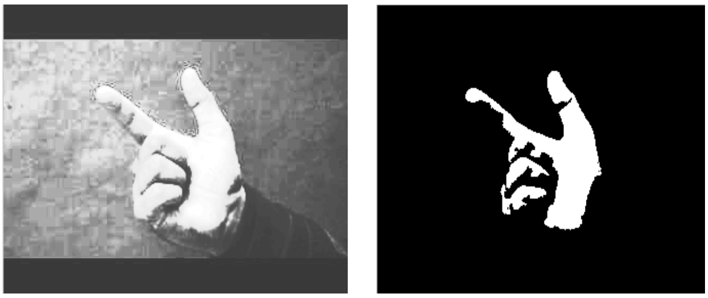

Figure 4: Segmentation results

Gray scale conversion is carried out on image frames followed by Otsu's thresholding which assumes that the image to be thresholded contains two classes of pixels, i.e. foreground and background. It calculates the optimum threshold by separating these two classes, so that their within class variance is minimum [16]. The within class variance is given as:

$$\sigma_w^2(t)= \omega_1(t)\, \sigma_1^2(t)+ \omega_2(t)\, \sigma_2^2(t) \qquad ...(1)$$

where

$\omega_1 = \sum_{i=1}^{T} P_i$

$\omega_2 = \sum_{i=T+1}^{L} P_i$

$P_i = n_i/N$, $(P_i \geq 0 \text{ and } \sum_{i=1}^{L} P_i = 1)$

Here, n_i is the number of pixels at *i*th gray level and *N* is the total number of pixels.

This method converts gray-level images to binary by turning the pixels below that threshold to zero and all pixels equals or exceeding that threshold to one. If g(x, y) is a threshold version of f(x, y) at some global threshold T, it can be defined as:

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) \geq T \\ 0, & otherwise \end{cases} \qquad ...(2)$$

C. Feature Extraction

Image segmentation gives us binary image that contains the hand shape representing a particular sign. For classification of this segmented gesture, there is need to extract certain features of the image that would be required for recognizing the sign. An effective shape descriptor can be a key component of image description, as shape is a fundamental property of any object. There are mainly two types of shape descriptors: region-based shape descriptors and contour-based shape descriptors.

In the proposed study, Hu's invariant moments are used for feature extraction. It is a popular type of contour-based shape descriptors that were firstly introduced by Hu in 1962 to the pattern recognition community [17]. A set of seven absolute orthogonal (i.e. rotation) moment invariants were derived from the results of the theory of algebraic invariants. These were computed by normalizing centralized moments upto order three. These can be used for scale, position and rotation invariant pattern identification.

- Translation invariance is achieved by computing moments that are normalized with respect to the centre of gravity so that the centre of mass of the distribution is at the origin (central moments).

- Size invariant moments are derived from algebraic invariants but these can be shown to be the result of simple size normalization.
- From the second and third order values of the normalized central moments, a set of seven invariant moments can be computed which are independent of rotation.

The seven moments are given as:

$$M1 = (\eta_{20} + \eta_{02}) \quad ...(3)$$

$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad ...(4)$$

$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad ...(5)$$

$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad ...(6)$$

$$M5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad ...(7)$$

$$M6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad ...(8)$$

Finally a skew invariant moment:

$$M7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad ...(9)$$

where η_{xy} is the normalized central moment.

First two invariant moments *M1* and *M2* are of order two, whereas the rest are third-order moments. The first one, *M1*, is analogous to the moment of inertia around the image's centroid, where the pixel's intensities are analogous to physical density. The last one, *M7*, enables it to distinguish mirror images of otherwise identical images.

D. Classification

Features calculated in the previous step are used as a basis for classification. In the proposed study, three layer Feed forward Neural Network is adopted for classification of signs [18], as neural network has emerged as a technology of preference for applications, like gestures identification, pattern recognition and system identification due to its simplicity and property of generalization. It has the ability to learn relationships from modeled data directly. At the same time, these are adaptive systems that are flexible in changing environment and meet real-time recognition constraints. Back propagation will be used for training of the network by adjusting weights between the elements. The proposed network is designed as follows:

TABLE I: DESCRIPTION OF PROPOSED NEURAL NETWORK

Number of input neurons:	7
Number of hidden neurons:	10
Number of output neurons:	10
Activation function:	Log Sigmoid

Seven features of image are fed as input to the network. So, input layer contains seven neurons. Output layer of network consists of ten neurons, as images are classified into ten classes, both for static and dynamic signs. Log sigmoid used as the transfer function for input and hidden layer is given as:

$$y(x) = \frac{1}{1 + e^{-x}} \quad ...(10)$$

Calculated features are used for training and testing of the proposed methodology. Two pairs of training and target vectors are created for static and dynamic signs.

Training vector: contains the feature vector. As seven features are considered for training and testing of 1000 images, the dimensions of feature vector for both pairs is 7×1000.

Target vector: contains the class vector to which the corresponding training vector belongs to. The images need to be classified into 10 classes. Therefore, the dimensions of target vector for each pair is 10×1000.

E. Performance Parameters

The selection of correct metrics in evaluating the performance of system is vital to the result and the validation of the system. The parameters are selected in such a way that effectiveness of the processes involved can be measured. Sensitivity, specificity and accuracy are chosen to evaluate performance of the proposed method.

Sensitivity: It relates to the test's ability to identify a condition correctly [19]. It measures the proportion of actual positives which are correctly identified as such, and is complementary to the false negative rate.

$$Sensitivity = \frac{TP}{TP + FN} \quad ...(11)$$

where TP is number of true positives (i.e. relevant items that are correctly identified as relevant) and FN is number of false negatives (i.e. relevant items that are incorrectly identified as irrelevant).

Specificity: It relates to the test's ability to exclude a condition correctly [19]. It measures the proportion of negatives which are correctly identified as such, and is complementary to the false positive rate.

$$Specificity = \frac{TN}{TN + FP} \quad ...(12)$$

where TN is number of true negatives (i.e. irrelevant items that are correctly identified as irrelevant) and FP is number of false positives (i.e. irrelevant items that we incorrectly identified as relevant).

Accuracy: It is the proportion of the test results that is true positive and true negatives among total number of cases [19].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad ...(13)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

Implementation of system is carried out using MATLAB® 2010b. The proposed methodology is tested for 20 signs, 10 static and 10 dynamic ones. 70% of the data in dataset is used for training and the remaining 30% for validation and testing.

The results for Static signs are shown in Table 2. It has been observed that the most of signs are correctly classified to their respective class.

TABLE II: RESULTS FOR STATIC SIGNS

Signs (class)	**No. of samples**	**True positives (*TP*)**	**False negatives (*FN*)**	**False positives (*FP*)**	**True negatives (*TN*)**
Flag	100	100	0	0	900
Marry	100	100	0	0	900
Aboard	100	100	0	0	900
Beside	100	100	0	0	900
All gone	100	100	0	1	899
Ascend	100	99	1	0	900
Middle	100	100	0	0	900
Hang	100	100	0	0	900
Anger	100	100	0	0	900
Moon	100	100	0	0	900

Therefore, the accuracy, sensitivity and specificity for all the signs is nearly equal to 100%. This is shown in Figure 5 below:

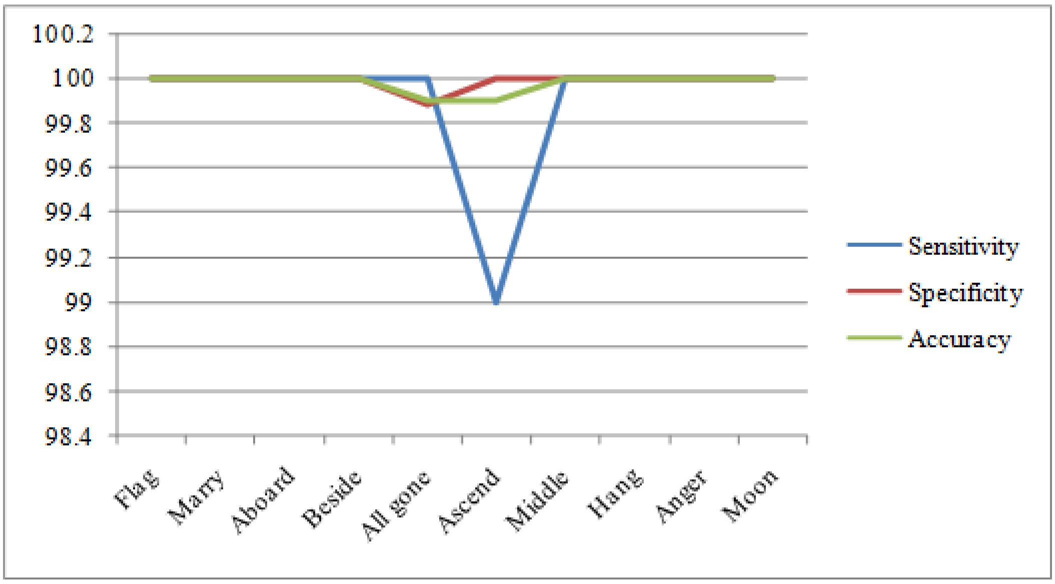

Figure 5: Calculated Performance Metrics for Static Signs

The result for Dynamic signs in Table 3 shows that system is able to correctly identify most of the signs. The performance drops slightly in the cases, when two or more signs share similar hand postures but have different direction of motion. The Hu's invariant moments give similar features for these similar signs. Therefore, the network does not get able to distinguish them.

TABLE III: RESULTS FOR DYNAMIC SIGNS

Signs (class)	**No. of samples**	**True positives (*TP*)**	**False negatives (*FN*)**	**False positives (*FP*)**	**True negatives (*TN*)**
Across	100	100	0	0	900
Above	100	100	0	0	900
Advance	100	100	0	0	900
All	100	100	0	1	899
Bag	100	100	0	0	900
Alone	100	100	0	0	900
Arise	100	77	23	0	900
Afraid	100	99	1	23	877
Yes	100	100	0	0	900
Bring	100	100	0	0	900

The corresponding values of accuracy, sensitivity and specificity for these signs are shown in Figure 6 below:

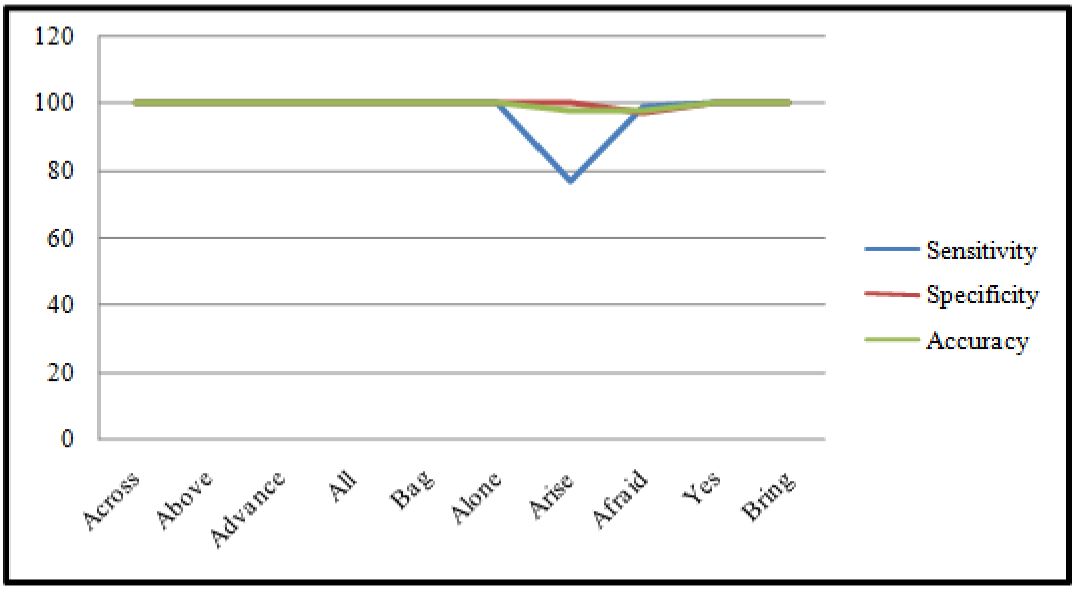

Figure 6: Calculated Performance Metrics for Dynamic Signs

Overall results of the system for both static and dynamic signs are shown in Table 4. The sensitivity slightly drops in case of dynamic signs. But the overall results are promising which makes the mark for the acceptability of proposed method.

TABLE IV: OVERALL RESULTS

	No. of signs	**Overall Sensitivity (%)**	**Overall Specificity (%)**	**Overall Accuracy (%)**
Static	10	99.9	99.98889	99.98
Dynamic	10	97.6	99.73157	99.51837

V. Conclusion

This paper presented a vision based approach for automatically recognizing static and dynamic hand gestures using artificial neural network. System was implemented in MATLAB® 2010b that calculates seven shape based features of segmented hand region to train the network. The use of Hu's invariant moments for feature extraction overcomes the challenges of position, size and rotation variations of the signs in images. Performance of the proposed system is evaluated in terms of Sensitivity, Specificity and Accuracy. The results show that the system is able to identify both static as well as dynamic signs.

In future work, some more features can be added for dynamic gestures that have similar shape but different direction of motion, to increase the accuracy.

ACKNOWLEDGEMENT

Thanks to Dr. Pavan Chakraborty, Ass. Professor at IIIT (Allahabad), for providing the dataset of ISL gestures and granting the permission to use it for completion of the research work.

REFERENCES

[1] "Verbal and Non-verbal." [Online]. Available: www.positive-parenting-skills.net/verbal and non-verbal communication. [Accessed: 26-Nov-2014].

[2] Kin fun li, E. Gill, and S. Lau, "A Web-Based Sign Language Translator Using 3D Video Processing," in *Proceedings of 14th International Conference on Network-Based Information Systems (NBIS)*, 2011, pp. 356–361.

[3] P. R. Futane and R. V. Dharaskar, "Video Gestures Identification And Recognition Using Fourier Descriptor And General Fuzzy MinMax Neural Network For Subset Of Indian Sign Language," in *Proceedings of 12th International Conference on Hybrid Intelligent Systems (HIS)*, 2012, pp. 525–530.

[4] V. Adithya, P. R. Vinod, and U. Gopalakrishnan, "Artificial Neural Network Based Method for Indian Sign Language Recognition," in *Proceedings of IEEE Conference on Information Technology*, 2013, pp. 1080–1085.

[5] F. M. Rahim, T. E. Mursalin, and N. Sultana, "Intelligent Sign Language Verification System – Using Image Processing , Clustering and Neural Network Concepts," *Int. J. Eng. Comput. Sci. Math.*, vol. 1, no. 1, pp. 43–56, 2010.

[6] N. Baranwal, N. Singh, and G. C. Nandi, "Indian Sign Language Gesture Recognition Using Discrete Wavelet Packet Transform," in *Proceedings of International Conference on Signal Propagation and Computer Technology (ICSPCT)*, 2014, pp. 573–577.

[7] K. Dixit and A. S. Jalal, "Automatic Indian Sign Language Recognition System," in *Proceedings of IEEE 3rd International Advance Computing Conference (IACC)*, 2013, pp. 883–887.

[8] D. K. Ghosh and S. Ari, "A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network," in *Proceedings of 8th International Conference on Information, Communications and Signal Processing (ICICS)*, 2011, pp. 1–5.

[9] M. E. Al-ahdal, N. Tahir, and U. T. Mara, "Review in Sign Language Recognition Systems," in *Proceedings of IEEE Symposium on computers & Informatics (ISCI)*, 2012, pp. 52–57.

[10] K. Dabre, "Machine Learning Model for Sign Language Interpretation using Webcam Images," in *Proceedings of IEEE Conference on Circuits , Systems , Communication and Infomation Technology Applications (CSCITA)*, 2014, pp. 317–321.

[11] A. Nandy, S. Mondal, J. S. Prasad, C. P, and G. C. Nandi, "Gesture based imitation learning for Human Robot Interaction," in *Proceedings of IEEE International Conference on Computer and Communication Technology (ICCCT)*, 2010, vol. 3749, pp. 712–717.

[12] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," in *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02)*, 2002, vol. 5, pp. 2204–2206.

[13] C. Oz and M. C. Leu, "Engineering Applications of Artificial Intelligence American Sign Language word recognition with a sensory glove using artificial neural networks," *Eng. Appl. Artif. Intell.*, vol. 24, no. 7, pp. 1204–1213, 2011.

[14] A. Nandy, S. Mondal, J. S. Prasad, P. Chakraborty, and G. C. Nandi, "Recognizing & Interpreting Indian Sign Language Gesture for Human Robot Interaction," in *Proceedings of IEEE International Conference on Computer and Communication Technology (ICCCT)*, 2010, pp. 712–717.

[15] Gonzalez and R. C. Wood, *Digital Image Processing*. Pearsons Education India, 2009.

[16] H. J. Vala and P. A. Baxi, "A Review on Otsu Image Segmentation Algorithm," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 2, no. 2, pp. 387–389, 2013.

[17] W.-H. Wong, W.-C. Siu, and K.-M. Lam, "Generation of invariants and their use for character recognition," *Pattern Recognit. Lett.*, vol. 16, no. 2, pp. 115–123, 1995.

[18] A. J. Maren, C. T. Harston, and R. M. Pap, *Handbook of Neural Computing Applications*. Academic Press, 2014.

[19] W. Zhu, N. Zeng, and N. Wang, "Sensitivity , Specificity , Accuracy , Associated Confidence Interval and ROC Analysis with Practical SAS ®," in *Proceedings of NESUG Health Care and Life Sciences*, 2010, pp. 1–9.