International Journal of Technology (2017) 3: 559-567
ISSN 2086-9614

HAND GESTURE RECOGNITION USING ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM AND K-NEAREST NEIGHBOR

Fifin Ayu Mufarroha[1], Fitri Utaminingrum[1*]

[1] *Computer Vision Research Group, Departement of Computer Science, Faculty of Computer Science, Universitas Brawijaya, Veteran Street 8 Malang, East Java 65145, Indonesia*

(Received: February 2016 / Revised: February 2017 / Accepted: April 2017)

ABSTRACT

The purpose of the study was to investigate hand gesture recognition. The hand gestures of American Sign Language are divided into three categories—namely, fingers gripped, fingers facing upward, and fingers facing sideways—using the adaptive network-based fuzzy inference system. The goal of the classification was to speed up the recognition process, since the process of recognizing the hand gesture takes a longer time. All pictures in all of the categories were recognized using K-nearest neighbor. The procedure involved taking real-time pictures without any gloves or censors. The findings of the study show that the best accuracy was obtained when the epochs score was 10. The proposed approach will result in more effective recognition in a short amount of time.

Keywords: Adaptive Network Based Fuzzy Inference System (ANFIS); American Sign Language (ASL); Hand gesture; K-nearest Neighbor (K-NN)

1. INTRODUCTION

Computer vision has been applied in various fields of science. One of the applications is in human–computer interaction. Communication is a process of conveying information between two individuals. Spoken communication allows individuals to deliver information or messages. However, transferring information is different for people with special needs especially deaf people. Sign language is a means of communication for people with special needs. Sign language comprises hand, arm, head, and face gestures as well as facial expressions. Hand gesture refers to different hand positions or movements, where each one has a distinctive meaning. In this study, hand gestures were identified and distinguished into letters in order to enable people with special needs to understand the information or message being conveyed. In this case, the key is how to distinguish each sign from things that change our orientation, so that we are able to use shape as a parameter (Panwar, 2012). In certain situations, some problems may arise during the process of recognition. In order to manage human activity recognition indoors, translation-based video and invariant scale are essential tools (Jalal et al., 2012). Hand gestures have various divisions, such as both static and dynamic movements, static movement, and dynamic movement (Ibraheem & Khan, 2012).

In general, deaf people do not have physical differences hearing people, the only difference is in how they communicate. However, people generally do not know and understand the sign language that deaf people use, and there tends to be a limited number of volunteers to help in translating sign language in social environments. Thus, communication between deaf and

*Corresponding author's email: f3_ningrum@yahoo.com, Tel. +0341-577-911, Fax. +0341-577-911
Permalink/DOI: https://doi.org/10.14716/ijtech.v8i3.3146

hearing people can be difficult and restricted. Some of them may know sign language from studying books, the Internet, or a forum assembled to learn sign language from proficient and experienced tutors. Not knowing and understanding sign language and different modes of communication creates gaps in social life. The above problems necessitate a solution that can overcome the widening gap between deaf people and hearing people. An application was made as a communication bridge between them, allowing hearing people to understand the communication used by deaf people. Further, this will allow deaf people to mingle in a social environment in such a way that social inequality will be resolved.

Soft computing is a set of measurement technique in the study for analyzing and modeling such complex problems. Soft computing was a combination of fuzzy logic, neuro computing, evolutionary computation, and probability computing in Dr. Lotfi A. Zadeh's first multidisciplinary study on soft data analysis (Zadeh, 2006). The adaptive network-based fuzzy inference system (ANFIS) is a combination of fuzzy logic and neural network. ANFIS has been designed for experiments aimed at identifying hand gesture recognition in Arabic (Al-Jarrah & Halawani, 2001). Furthermore, its established ability to identify brain tumors in humans is evidence that ANFIS is a sophisticated platform for multi-object classification (Deshmukh & Khule, 2014). Another benefit of ANFIS is TIMIT speech database recognition (Silarbi et al., 2014).

The purpose of the study was to recognize 26 types of real-time American Sign Language (ASL) using a static hand and capturing hand gestures using a web camera. The user did not use gloves or wrist censors. A different approach was implemented in the recognition. The researchers used the ANFIS learning algorithm, whose function is to divide images into particular groups. The following step is to recognize the hand signals from the groups using K-nearest neighbor (K-NN).

2. THE PROPOSED SYTEM DESIGN

The researchers developed a system to recognize ASL. Figure 1 shows a system of recognition consisting of several processes—namely, image capturing, pre-processing, grouping, classification and output.

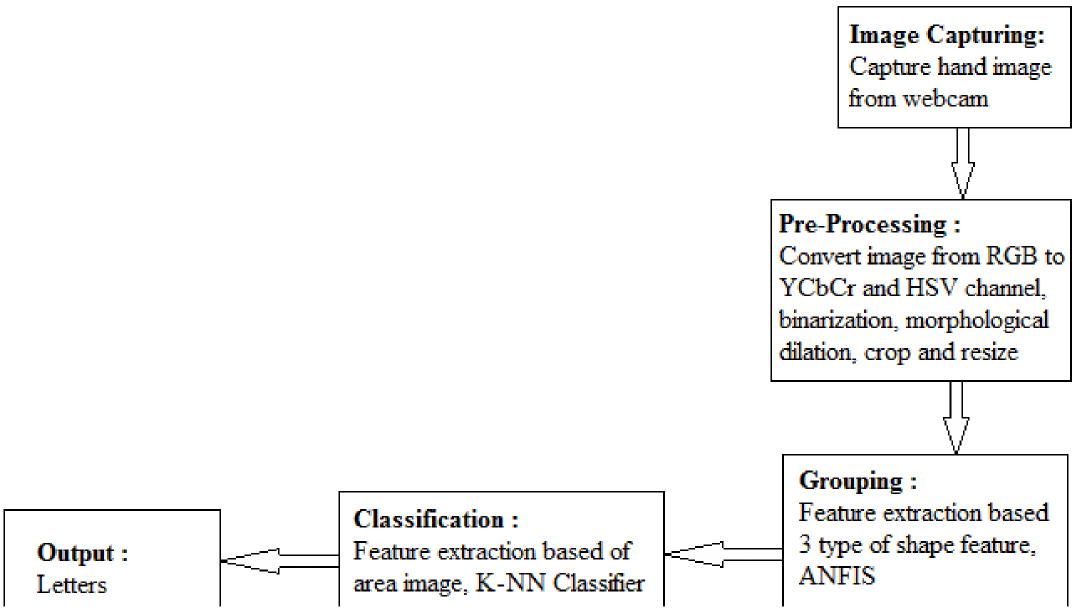


Figure 1 Proposed method

3. PRE-PROCESSING

Figure 2 displays all 26 hand gestures in ASL that were used in the data collection and experiment. During learning and experimentation, the researchers used five pieces of data per hand gesture. Thus, the data comprised 130 (or 5×26) hand gestures. At the pre-processing stage, the researchers obtained natural images of the hand gestures, separated from the background.

In the process of picture-taking, the researchers got the same measurement or size of the hand gestures as what the web camera captured. Some noise and empty space mix together when a photo of an object is taken using a web camera. The researchers used automatic editing (cropping) to select hand objects from the photos. In the process, the researchers converted RGB images into YCbYCr and HSV channels. The researchers then detected skin and changed the photos into binary images. Some noise can be resolved using morphological dilation; however, an adaptive window-based quick filter can also be used to get rid of impulse noise (Fitri et al., 2015). Finally, the researchers cropped and converted the size of the images into 300×400 pixels. An example of the binary images is shown in Figure 3.

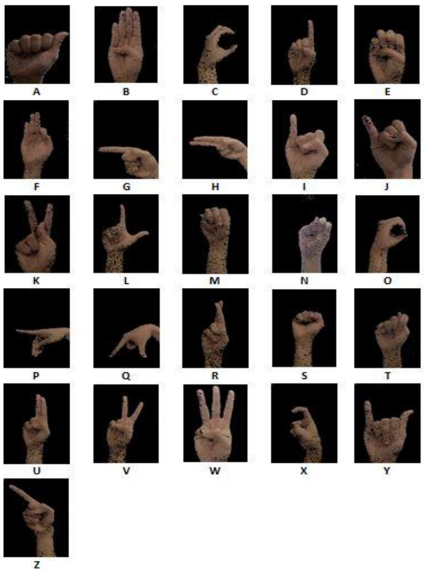


Figure 2 American Sign Language

Figure 3 Binary Image

4. ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM) METHOD

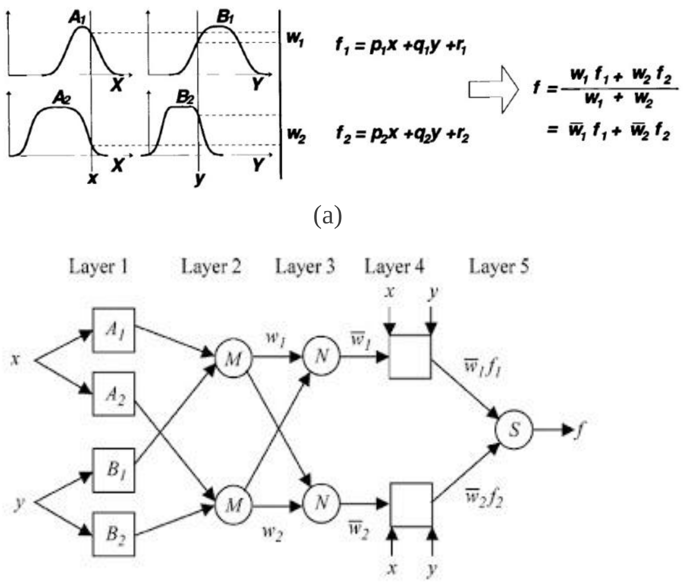
4.1. Concept and Structure

The fuzzy system provides a neuro-system with a highly structured platform and thought and analysis IF–THEN fuzzy rules, whereas the neuro-system provides the fuzzy system with learning ability. A couple of input and output data were used to establish learning procedures that optimized the parameters. Figure 4 shows the Takagi-Sugeno-Kang (TSK) model of fuzzy elaboration. For clarify, the researchers considered X and Y as the inputs of the fuzzy inference system and Z as the output (Jang, 1993). The illustration of the process is shown in Figure 4c. The model of the fuzzy inference TSK system was the first system used in the fuzzy inference system with architectural rules, as shown in Equations 1 and 2:

$$Rule\ 1: if\ x\ is\ A1\ and\ y\ is\ B1, then\ f1 = p1x + q1y + r1 \tag{1}$$

$$Rule\ 2: if\ x\ is\ A2\ and\ y\ is\ B2, then\ f2 = p2x + q2y + r2 \tag{2}$$

x and y are the fuzzy inputs. $A1$ and $A2$ are membership functions, while $f1$ and $f2$ are the outputs with $p1, q1, r1, p2, q2$ and $r2$ as the parameters of consequence.



(a)

(b)

Figure 4 Takagi-Sugeno-Kang Model: (a) Type-3 Fuzzy Reasoning; (b) Type-3 ANFIS (Equivalent ANFIS)

Layer 1: Every node i in the layer is an adaptive node. In the layer, x is the input to node i, and the function of node i is the Gaussian function to measure Ai membership from the activation function node as shown in the following Equation 3:

$$O_i^1 = \mu A_i(x) = e^{-((x-c_i/\sigma_i))} \tag{3}$$

Layer 2: The degree of activation (firing strength) of each fuzzy rule is stated on each output node. The output is a multiplication of all the inputs that enter this layer.

$$O_i^2 = w_i = \mu A_i(x) \times \mu B_i(y) \tag{4}$$

Layer 3: Normalized firing strength. The i-th node calculates the ratio of the i-th rule's firing strength to the sum of all rules' firing strengths, as shown in Equation 5:

$$O_i^4 = \bar{w} = \frac{w_i}{w_1+w_2} \tag{5}$$

Layer 4: Every node in the layer is an adaptive knot. Output from layer 3 is $\bar{w}_i$, and the parameter set is $\{p_i, q_i, r_i\}$. The parameters in the layer are called consequence parameters, and Equation 6 is as follows:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{6}$$

Layer 5: Sum up all inputs to measure the output of the fuzzy system, as shown in Equation 7:

$$O_i^5 = f = \sum_{i=1}^{4} \bar{w}_i f_i = \frac{\sum_{i=1}^{4} w_i f_i}{\sum_{i=1}^{4} w_i} \tag{7}$$

4.2. Learning Algorithm to Determine the Group with ANFIS

Hand gesture recognition in the proposed method meant dividing the hand gestures into three groups. There were 26 hand signals that followed a pattern of hand gestures. To determine the pattern of the hand gestures, the researchers decided to distinguish groups to distribute the same patterns. The basis of the group division was the shape of every hand movement. The division of the hand gestures is presented in Table 1.

Table 1 Group division based on hand gestures

Group	Alphabet
Fingers gripped	A, E, M, N, O, S, T
Fingers upward	B, C, D, F, I, K, L, R, U, V, W, X, Y
Fingers sideways	G, H, J, P, Q, Z

Given the sophistication of the ANFIS learning algorithm, the researchers decided to ease the extraction feature. The three extraction features used as the inputs for the ANFIS method are as follows.

Slimness: The feature is defined as the ratio between the length and width of a hand (Wu et al., 2006). L_p is the length of the hand image and W_p is the width of the hand. See Equation 8 as follows:

$$slimness = \frac{L_p}{W_p} \tag{8}$$

Roundness: This is defined in Equation 9:

$$roundness = \frac{4\pi A}{P^2} \tag{9}$$

A is the area of the hand image and *P* is the circumference of the hand image.

Rectangularity: This feature illustrates the similarity of the hand technique with a rectangular-shaped box (Singh et al., 2010) and is defined as in the following Equation 10:

$$rectangularity = \frac{L_p W_p}{A} \tag{10}$$

L_p and W_p are the length and width, respectively, of the hand image, while *A* is the size of the hand image.

Based on the three types of image extraction features, the hand images were classified into three groups in ANFIS. The researchers approached the membership function type as Gaussian, and the epochs were 5, 10, and 100, while the set of the membership function was 3.

5. THE RECOGNITION SYSTEM

In the recognition process, the pattern of hand signals was categorized in the previous stage, where the images classified by ANFIS were to be extracted by looking for areas in the images (Naik & Metkewar, 2015). In order to measure how features an image had, each image was divided into 10×10 areas. The feature was obtained by the accumulation of pixels in each area. The total number of features in the process of feature extraction was 100, as illustrated in Figure 5.

The K-NN algorithm has been used extensively in various studies. The classification algorithm is simple and easy to apply as a learning utility algorithm (Bishop, 2006), and K-NN is a very simple classifier that works well with basic recognition problems. In this study, we divided the hand gestures into groups using the ANFIS method. We chose this method to ease the computation time. K-NN is carried out using a set of data obtained at feature extraction and a target class that is going to be compared to the score of the features of the data. The goal of the process is to find the shortest distance between the features of the data in the training process and those of the data in the testing process. K-NN obtains classes on the basis of the total mean of the classes close to the K spot. The shortest matrix distance is measured using the Euclidean distance.

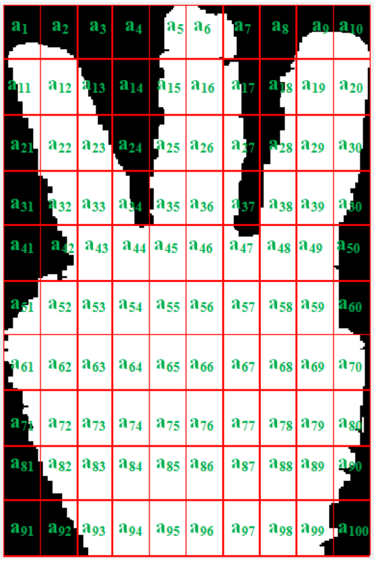


Figure 5 The 10×10 area of features

6. RESULTS

This section presents the findings of the study. The algorithm was determined for hand gesture recognition.

6.1. The Results of ANFIS Simulation

The researchers used 130 pieces of testing data were used and three test scenarios were carried out. The mean squared error (MSE) measures the mean of the squared errors. Smaller MSE scores guarantee a better performance. p_n and q_n represent the network outputs and the score measured by the n element, and k refers to the number of observations, as shown in the following Equation 11 (Ahmed & Shah, 2015):

$$MSE = \frac{1}{k}\sum_{n=1}^{k}(p_n - q_n)^2 \tag{11}$$

The data variance in the distribution of the ANFIS training is illustrated in Figure 6, while the MSE results are shown in Table 2.

Table 2 Results of grouping using ANFIS

Scenario	# of epochs	Phase	MSE	# of samples	Grouping samples	EFF
1	5	Training	2.49	130	112	86.15%
		Testing	3.07	130	95	73.08%
2	10	Training	2.22	130	113	86.92%
		Testing	2.19	130	105	80.77%
3	100	Training	2.49	130	112	86.15%
		Testing	3.33	130	92	70.77%

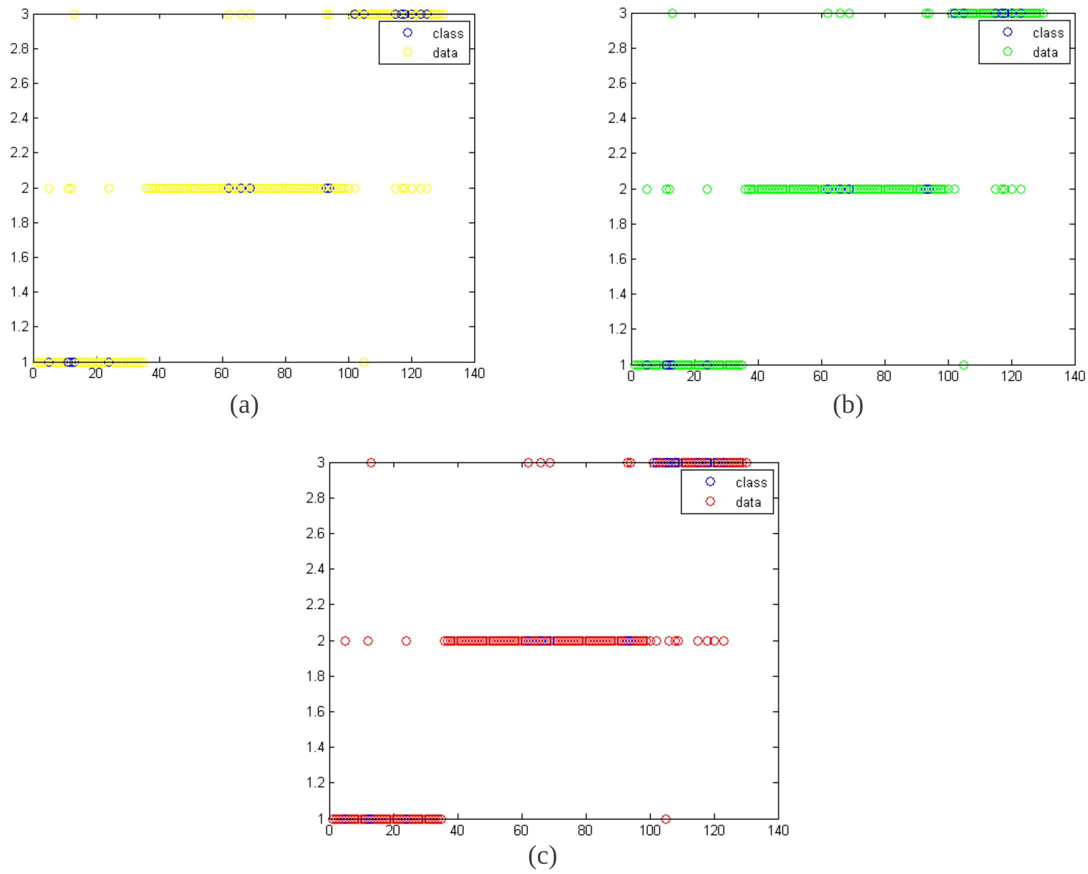


Figure 6 Number of ANFIS Epochs: (a) 5 Epochs; (b) 10 Epochs; (c) 100 Epochs

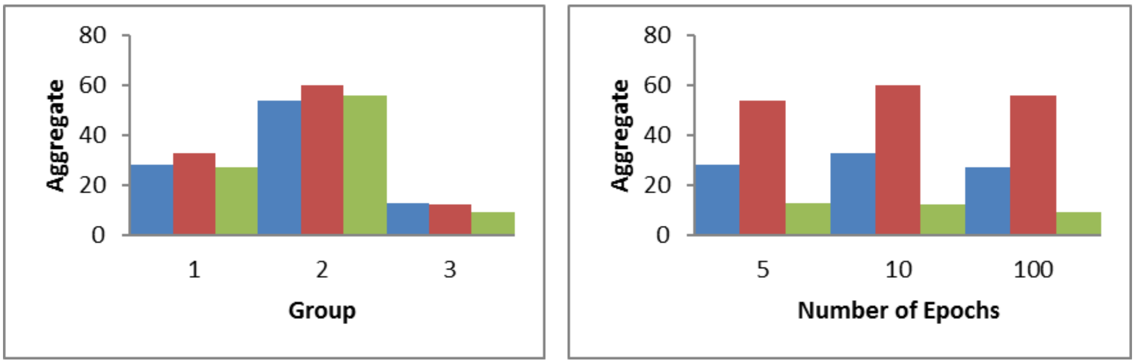
6.2. The Results of Recognition

After obtaining the results of grouping or classification using ANFIS with different numbers of epochs, the images were then translated into text using K-NN.

Table 3. Recognition results using K-NN

Scenario	# of epochs	Group	# of samples	Recognition sample	Recognition result	Aggregate
1	5	Fingers gripped	35	27	77.14%	69.30%
		Fingers upward	65	50	76.92%	
		Fingers sideways	30	13	43.33%	
2	10	Fingers gripped	35	33	94.29%	80.77%
		Fingers upward	65	60	92.31%	
		Fingers sideways	30	12	40.00%	
3	100	Fingers gripped	35	27	77.14%	70.77%
		Fingers upward	65	56	77.14%	
		Fingers sideways	30	9	30.00%	

Table 3 shows that the best results were obtained in the test where there were 10 epochs, as the accuracy was 80.77%. Sufficient grouping resulted in a sufficient recognition process. The comparison of the recognition results using all numbers of epochs is depicted in Figure 7. Furthermore, some previous studies using the K-NN algorithm are listed in Table 4.



(a) (b)

Figure 7 Results of recognition: (a) Each Group; (b) Each Number of Epochs

Table 4 Comparison between previous studies using the K-NN algorithm and the proposed method

No	Circumstances	Type of Classifier	Accuracy	Reference
1	SHIFT with clear Background	K-NN	68.00%	(Rekha, et.al., 2011)
	SHIFT with slightly cluttered background	K-NN	60.00%	
	SHURF	K-NN	84.00%	
2	PCA (train data)	K-NN	75.54 %	(Kotha, et.al., 2015)
	LDA (train data)	K-NN	24.50 %	
3	ANFIS with 5 epochs	K-NN	69.30%	Proposed method
	ANFIS with 10 epochs	K-NN	80.77%	
	ANFIS with 100 epochs	K-NN	70.77%	

7. CONCLUSION

The purpose of this study was to propose a new approach for hand gesture recognition. The approach involves creating groups on the basis of the hand patterns using the ANFIS method. The researchers classified 26 hand gestures into 3 categories—fingers gripped, fingers facing upward, and fingers facing sideways. Then the hand gesture recognition was carried out with the groups using K-NN classifications. The findings show that the best accuracy was obtained when the number of epochs was 10, with 80.77% accuracy. For further study, the researchers have made an improvement in their studies by using the feature extraction differences in the hope of obtaining more accurate results in the future.

8. REFERENCES

Ahmed, A.A.M., Shah, S.M.A., 2015. Application of Adaptive Neuro Fuzzy Inference System (ANFIS) to Estimate the Biochemical Oxygen Demand (BOD) of Surma River. *Journal of King Saud University – Engineering Sciences*, In Press, Corrected Proof, pp. 1–7

Al-Jarrah, O., Halawani, A., 2001. Recognition of Gesture in Arabic Sign Language using Neuro–Fuzzy System. *Artificial Intelligence*, Volume 133, pp. 117–138

Bishop, C.M., 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer New York Inc. Secaucus

Deshmukh, R.J., Khule, R.S., 2014. Brain Tumor Detection using Artificial Neural Network Fuzzy Inference System (ANFIS). *International Journal of Computer Applications Technology and Research*, Volume 3(3), pp. 150–154

Fitri, U., Keiichi, U., Gou, K., 2015. Speedy Filters for Removing Impulse Noise Based on an Adaptive Window Observation. *AEU-International Journal of Electronics and Communications*, Volume 69(1), pp. 95–100

Ibraheem, N.A., Khan, R.Z., 2012. Vision Based Gesture Recognition using Neural Networks Approaches: A Review. *International Journal of Human Computer Interaction (IJHCI)*, Volume 3(1), pp. 1–14

Jalal, A., Uddin, M.Z., Kim, T.S., 2012. Depth Video-based Human Activity Recognition System using Translation and Scaling Invariant Features for Life Logging at Smart Home. *IEEE Transactions on Consumer Electronics*, Volume 58(3), pp. 863–871

Jang, J.S.R., 1993. ANFIS: Adaptive Network-based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 23(3), pp. 665–685

Kotha, S.K., Pinjala, J., Kasoju, K., Pothineni, M., 2015. Gesture Recognition System. *IJRET: Internaional Journal of Research in Engineering and Technology*, Volume 4(5), pp. 99–104

Naik, S., Metkewar, P., 2015. Recognizing Offline Handwritten Mathematical Expression (ME) based on a Predictive Approach of Segmentation using K-NN Classifier. *International Journal of Technology*, Volume 3, pp. 345–354

Panwar, M., 2012. Hand Gesture Recognition Based on Shape Parameters. *In:* 2012 International Conference on Computing, Communication and Applications (ICCCA), Volume 1(6), pp. 22–24

Rekha, J., Bhattacharya, J., Majumder, S., 2011. Hand Gesture Recognition for Sign Language: A New Hybrid Approach. *In:* The 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), pp. 80–86

Silarbi, S., Abderrahmane, B., Benyettou, A., 2014. Adaptive Network Based Fuzzy Inference System for Speech Recognition through Subtractive Clustering. *International Journal of Artificial Intelligence & Application (IJAIA)*, Volume 5(6), pp. 43–52

Singh, K., Gupta, I., Gupta, S., 2010. SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition*. Volume 3(4). pp. 67–78

Wu, Q., Changle, Z., Chaonan, W., 2006. Feature Extraction and Automatic Recognition of Plant Leaf using Artificial Neural Network. *Advances in Artificial Intelligence*, Volume 3, pp. 5–12

Zadeh, L.A., 2006. *Soft Computing*. Available online at www.cs.berkeley.edu/~zadeh