

Static Hand Gesture Recognition for Human Computer Interaction

Hamid A. Jalab

Faculty of Computer Science and Information Technology, University Malaya,
50603 Kuala Lumpur, Malaysia

Abstract: This study presents a novel algorithm to recognize a set of static hand gestures for the Human-Computer Interaction (HCI), based on hand segmentation using both wavelet network for images feature extraction, and supervised feed-forward neural network with back propagation training algorithm for recognition. One hundred and twenty hand gesture images were used for training and 60 for testing. The best classification rate of 97% was obtained for the testing set.

Key words: Gesture recognition, human-computer interaction, wavelet network, neural network

INTRODUCTION

Of late, the research world has been actively involved in, inventing devices and methods to enhance the spontaneity of human computer interaction. The gesture recognition is one of the most potential methods in this domain. A simple definition of the term gesture is suggestive movement of bodily parts such as fingers, arms etc, which convey some information. According to Kurtenbach and Hulteen (1990), "A gesture is a motion of the body that conveys information. Waving the hands is a gesture that suggests "good bye". However, the movement of body parts that does not convey specific meaning is not considered as gesture, such as pressing key in computer key board. Even though gestures can originate from any bodily movement, generally it originates from the movement of face or hand. Presently researchers focus more on emotion recognition in the field of face and hand gesture recognition. A lot of techniques and approaches have been proposed, which uses cameras and computer vision algorithms to construe sign language. Nevertheless recognizing posture, way of walking and manners is also active topic of gesture recognition.

The primary aim of hand gesture recognition is to build a system, which will enable their device control to understand specific hand gestures. Gesture recognition was first proposed by Myron W. Krueger as a new form of interaction between human and computer in the middle of seventies (Dong *et al.*, 1998). It has become a very significant research area with the swift advancements in computer hardware and vision systems. The key problem in gesture interaction is how to make hand gestures understood by computers. A lot of techniques related hand gesture recognition has been proposed recently. These techniques are bifurcated into data-glove based

and vision based approaches (Garg *et al.*, 2009). In the data glove-based gesture systems the user needs to wear burdensome accessories, which are generally connected to a computer (Chen *et al.*, 2003). On the other hand, the vision-based gesture recognition technique does not burden the users with the loads cables and special hardware. Instead it needs only a camera. Therefore the vision-based gesture recognition techniques are capable of understanding the natural humans-computer interaction devoid of any additional devices. At the same time, this leverage imposes a challenge to these systems, as they need consistent background, indifferent lighting, person and camera independent to accomplish coincident performance. Furthermore, these systems must be optimized to be accurate and robust.

Hand gestures fall into two categories such as: Macro gestures and micro gestures. The former represents various positions of the hands associated to the human body (dynamic gestures). Whereas, the micro gestures represent the relative position of the fingers of the hand (static gestures). A lot of algorithms to identify hand gestures have been proposed in a research, Hasanuzzaman *et al.* (2004) have presented a real-time hand gesture recognition system to detect micro hand gestures, using skin color segmentation and multiple-feature based template-matching techniques. Recently, Murthy and Jadon (2010) have proposed a simple vision based gesture recognition system, which uses a web camera to detect hand, count fingers and find the direction in the finger is pointed. Fang *et al.* (2007), proposed scale space feature detection such as Blob and Ridge detection to recognize certain hand gestures. Wu and Balakrishnan (2003), published a work using hand gestural interaction, in this work they used a touch surface to aid the gesture recognition. Kim and Fellner (2004), used marked fingertips and infrared light to track

hand motion and recognize gestures, they applied their work to 3D object manipulation and deformation. Malik and Laszlo (2004) and Malik *et al.* (2005) used hand gestures over a tabletop as a two-hand input device for large displays from a distance. They consider fingertips and gesture recognition as two completely distinct processing steps. Lee and Hong (2010) proposed a real-time hand gesture recognition system based on the difference image entropy obtained using a stereo camera. Dardas and Petriu (2011), presented a real time system, which includes detecting and tracking bare hand in cluttered background using skin detection and hand postures contours comparison algorithm after face subtraction and recognizing hand gestures using Principle Components Analysis (PCA).

PROPOSED WORK

The aim of this study was to propose a vision based hand gesture recognition algorithm using both wavelet network for images feature extraction and supervised feed-forward neural network with back training algorithm for classifying various hand gestures. The proposed system of hand gesture recognition consists of three basic stages: preprocessing, feature extraction using wavelet network and classification using neural network.

The primary goal of the preprocessing stage is to ensure uniform inputs to the classification network. This must include the following steps which are inspired by Murthy and Jadon (2010):

- Read the gesture color image, then convert the RGB image into grayscale as shown in Fig. 1b
- Hand segmentation to isolate the foreground (hand gesture) from the background. The absolute difference between the hand gesture and the background used to isolate the hand gesture from the background as shown in Fig. 1c

WAVELET NETWORK FEATURES EXTRACTION

Wavelet Transform (WT) has been introduced rather recently in mathematics, even through the essential ideas

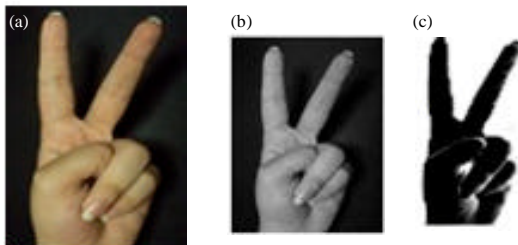


Fig. 1: Gesture image, (a) Color image, (b) Grayscale image and (c) Image with uniform background

that lead to this development have been around for a longer period of time (De Castro Fernandez and Rojas, 2002; Gu *et al.*, 2011). Wavelets are mathematical functions that cut up data into different frequency components and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes (Lekutai, 1997; Gu *et al.*, 2011).

Wavelet network algorithm: A new technique to extract features from gesture image using wavelet network has been proposed. This technique uses special mother wavelet $\Psi_{a,b}(\omega)$ as activation function for Artificial Neural Network (ANN) instead of the traditional activation function (Sharma and Agarwal, 2012). The wavelet network architecture shown in Fig. 2. Approximates any desired signal y by generalizing a linear combination of a set of daughter wavelets, where the daughter wavelets are generated by dilation a and translation b , from a mother wavelet. The selection of mother wavelet depends on the type of signal. If the signal represents function of two variables like images, then it will require two variable mother functions:

$$\Psi_{a,b}(t_1, t_2) = \Psi\left(\frac{t_1 - b}{a}, \frac{t_2 - b}{a}\right) \tag{1}$$

Where:

a: Dilation factor, with $a > 0$

b: Translation factor

The approximated signal of the network y can be represented by:

$$y(t) = u(t) \times \sum_{k=1}^K w_k \Psi_{a_k, b_k} \tag{2}$$

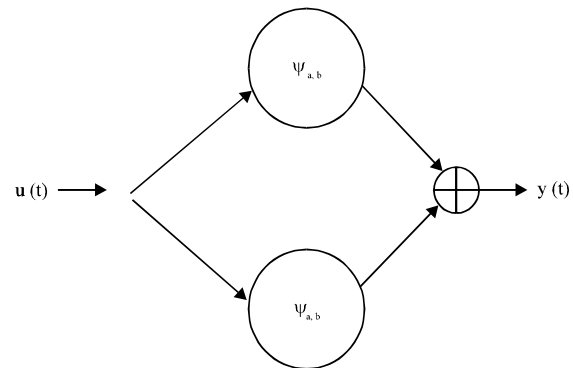


Fig. 2: Adaptive wavelet network structure

where, $y(t)$ represents approximated image, $u(t)$ is the input gesture image that we want to extract feature from it, K is a number of windowed wavelets, w_k is the weight coefficients (features coefficients) and $\Psi_{a,b}$ represents multivariable mother function.

Feature extraction: The feature extraction used to derive the suitable possible features from the segmentation region in order to use it to recognize the different input gestures. The wavelet network parameters w_k , a_k and b_k in multivariable functions can be optimized in the Least Mean Square algorithm (LMS) by minimizing a cost function or the energy function E , over all dimensions of function. Thus by denoting (Hasan, 2004; Melam and Amar, 2010; Muzhou and Xuli, 2010; Goswami and Chan, 2011):

$$e(i, j) = u(i, j) - y(i, j) \quad (3)$$

The energy function is defined by:

$$E = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N e^2(i, j) \quad (4)$$

$$E = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (u(i, j) - y(i, j))^2 \quad (5)$$

where, M and N are the dimensions of function.

To minimize E , method of steepest descent be used, which requires the gradients:

$$\frac{\partial E}{\partial w_k}, \frac{\partial E}{\partial a_k} \text{ and } \frac{\partial E}{\partial b_k}$$

for updating the incremental changes to each particular parameter w_k , a_k and b_k , respectively. The gradient of E can be define as follows:

$$\frac{\partial E}{\partial w_k} = -\frac{2}{M \times N} \sum_{i=1}^M \sum_{j=1}^N e(i, j) y(i, j) \Psi_k(\tau_1, \tau_2) \quad (6)$$

$$\frac{\partial E}{\partial b_k} = -\frac{2}{M \times N} \sum_{i=1}^M \sum_{j=1}^N e(i, j) y(i, j) w_k \frac{\partial \Psi_k(\tau_1, \tau_2)}{\partial b_k} \quad (7)$$

$$\frac{\partial E}{\partial a_k} = -\frac{2}{M \times N} \sum_{i=1}^M \sum_{j=1}^N e(i, j) y(i, j) w_k \frac{\partial \Psi_k(\tau_1, \tau_2)}{\partial a_k} \quad (8)$$

$$\tau_1 = \frac{i - b_k}{a_k} \quad (9)$$

$$\tau_2 = \frac{j - b_k}{a_k} \quad (10)$$

The incremental changes of each coefficient are simply the negative of their gradients:

$$\Delta w = -\frac{\partial E}{\partial w} \quad (11)$$

$$\Delta b = -\frac{\partial E}{\partial b} \quad (12)$$

$$\Delta a = -\frac{\partial E}{\partial a} \quad (13)$$

Thus each coefficient w , b and a of the network is updated in accordance to the following rules:

$$w(n+1) = w(n) + \mu_w \Delta w \quad (14)$$

$$b(n+1) = b(n) + \mu_b \Delta b \quad (15)$$

$$a(n+1) = a(n) + \mu_a \Delta a \quad (16)$$

where, μ is the fixed learning rate parameter, typically between (10 to 100).

Wavelet network feature extraction algorithm is composed of three main parts:

- Resize the input grayscale gesture images to 64×64 pixels, in order to decrease the processing time of wavelet network
- Randomly choose the initial parameters of wavelet network (dilation, number of wavelons and the learning rates for wavelet network). The learning rate parameters for weights, dilations and translations are fixed at 90 (typically between 10 to 100). Wavelet Network has 128 wavelons and the desired error is chosen equal to 0.05
- Store the final w_k wavelet network parameter (1×128) as features of gesture image to be used as features for classifications. The final feature vectors (120×128) that obtained from all 120 hand gesture images

NEURAL NETWORK

The design of Artificial Neural Network (ANN) has been inspired by the biological research on how the human's brain works. In an effect to model certain capabilities of the brain, Warren McCulloch and Walter Pitts established a simplified model of a biological neuron in 1943 called the McCulloch-Pitts model consisting of multiple inputs and one output

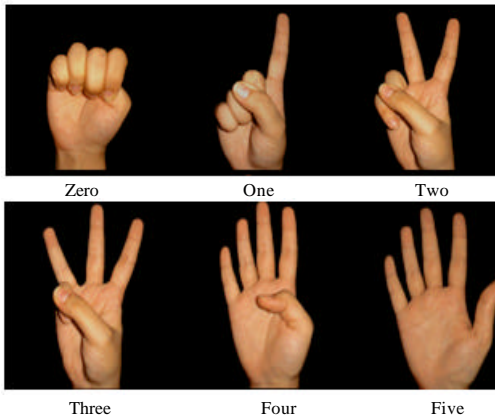


Fig. 3: Six hand gesture images classes

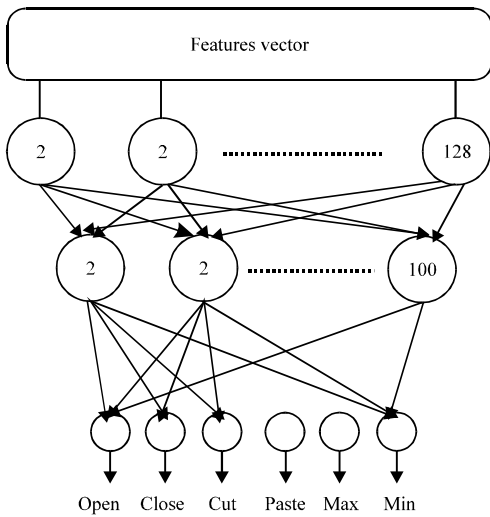


Fig. 4: Neural network classifier

Table 1: Target vector for hand gesture classes

Target	t1	t2	t3	t4	t5	t6
Zero	1	0	0	0	0	0
One	0	1	0	0	0	0
Two	0	0	1	0	0	0
Three	0	0	0	1	0	0
Four	0	0	0	0	1	0
Five	0	0	0	0	0	1

(Khanale and Chitnis, 2011). ANNs have been applied to a variety of real world classification tasks in industry, business and science (Taqa and Jalab, 2010; Bash, 2011; Yedjour *et al.*, 2011; Sharma and Agarwal, 2012).

Neural network based classifier: The last stage of the proposed system is the classification. We have used neural networks as they are most suitable solution for image recognition or sign classification (Al-Bashish *et al.*, 2011). The right choice of the classification algorithm to be used in a gesture recognition system is highly

dependent on the properties and the format of the features that represent the gesture image. In this work a standard back-propagation neural network is used to classify gestures. The network consists of three layers; the first layer consists of neurons that are responsible for input feature vectors into the neural network. The second layer is a hidden layer. This layer allows neural network to perform the error reduction necessary to successfully achieve the desired output. The final layer is the output layer which is determined by the size of the set of desired outputs, which represent the recognized gesture image. Each possible output being represented by a separate neuron. There are six outputs from neural network, each output represents index for one of the six hand gesture images classes, which are shown in Fig. 3. The neural network structure is shown in Fig. 4.

Training phase: The ANN is trained to classify hand gesture features. The hand gesture images dataset contains 120 images for training and 60 images for testing, captured by personal digital camera under natural conditions. All the images in image dataset are divided into 6 classes (Table 1). The feature vectors (120×128) that obtained from the images.

In the training phase of the ANN, the weight matrices between the input and the hidden and output layers are initialized with random values. After repeatedly presenting features of the input samples and desired values. The output data are compared with the desired values and the errors are computed. This pattern is repeated until the error rate of the output layer reaches a minimum value. This process is then repeated for the next input value, until all values of the input have been processed. The activation function used is binary-sigmoid. The value of this function ranges between 0 and 1. Whereas, the output layer neuron is estimated using the activation function that features the linear transfer function. The training algorithm used is gradient descent with momentum back propagation. Features are extracted by wavelet network and entered as training input data into the ANN. The quality of the training sets that enters into the network determines how well the classifier works.

Figure 5 shows the performance of the trained network, determined by performing a validation performance analysis between the mean squared error and the number of iterations. The best validation performance is 0.059332 at iteration number 11. The result is reasonable because the final mean-square error is small and no significant over fitting has occurred.

The regression analysis which is shown in Fig. 6, returns the correlation coefficient R. This coefficient equals to 0.88695 between the output and the target for training; thus, both output and target are very close, which indicates good fit.

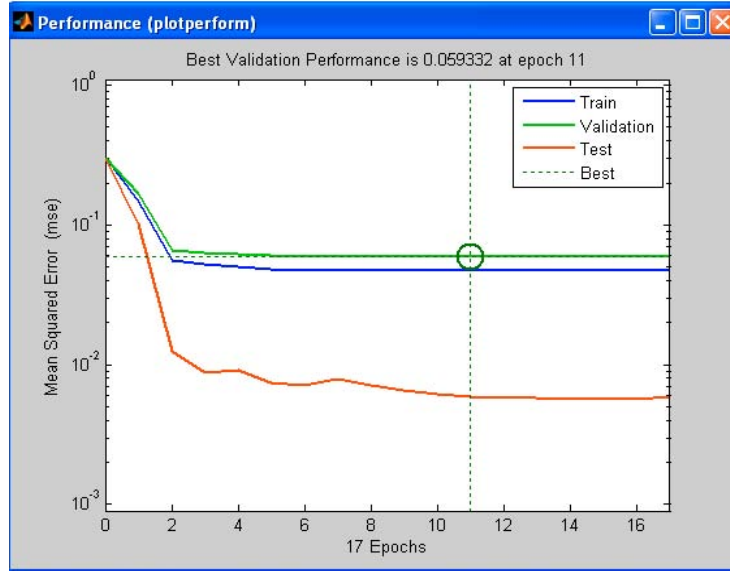


Fig. 5: Neural network training performance

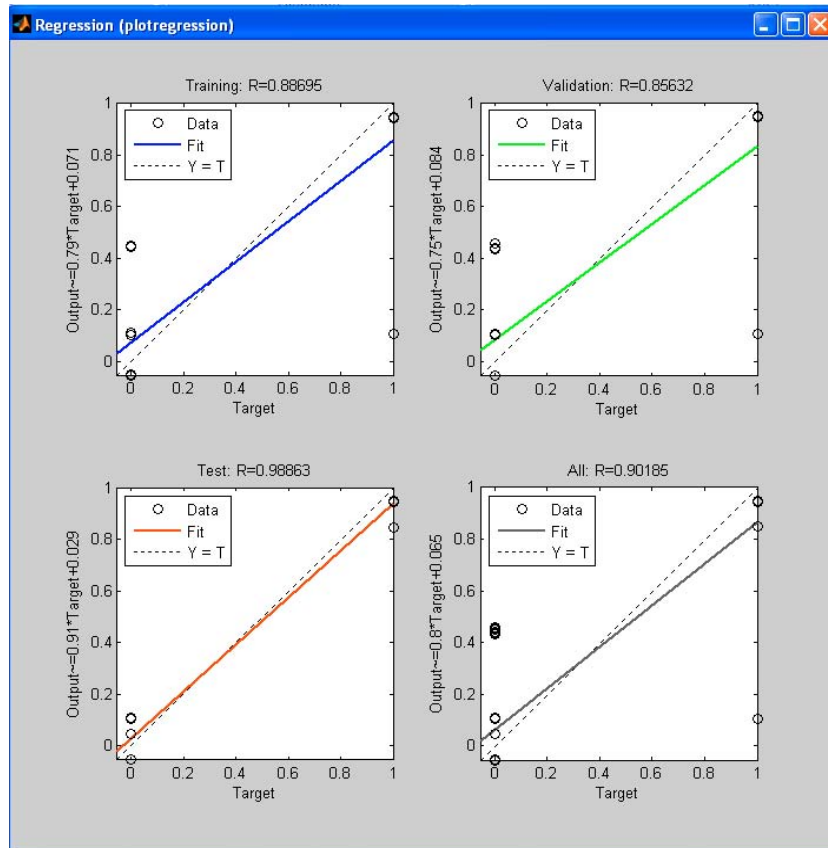


Fig. 6: Regressions analysis

Testing phase: In this phase, the features are extracted in the same manner as in training phase. In this phase, depending on training data, 60 hand gesture images are used to test the proposed system with different amount of Gaussian noise. All the testing images were divided into 6 classes.

RESULTS AND DISCUSSION

The experimental results are presented to show the effectiveness of the proposed system. Our hand gesture recognition system was carried out on a 2.33 GHz Intel (R) Core TM 2Duo CPU 2 GB RAM on Windows 7 platform using MATLAB R2010a. Table 2 shows the classifying network results for one hand gesture image.

This result shows that, the neural network works probably and absolutely no errors found in the outputs. The highest value of the neural network outputs represent the recognized gesture image. From Table 2, the output value which is equal to 0.548, represent the highest value in the classification result for one test hand gesture image. This image represent class (Two), while the system will ignore the other output values. We have achieved 97% recognition rate with our captured data. Due to noise addition to the gesture images false positives occurred. Wavelet network plays important role in recognition process. It succeeded to extract features from hand gesture image. Table 3, compares proposed method with other static hand gesture systems (Murthy and Jadon, 2010; Palaniappan *et al.*, 2010; Steinberg *et al.*, 2010) using different dataset images.

Table 2: Neural network output for hand gesture image

Class name	Neural network outputs
Zero	0.00224
One	0.000454
Two	0.548
Three	0.0326
Four	0.00686
Five	0.002543

Table 3: Comparison of classification methods

Classification rate (%)	Classification method used	
	Classification method used	Reference
97	Neural network	Proposed method
89	Neural network	Murthy and Jadon (2010)
97.19	Neural network	(Palaniappan <i>et al.</i> (2010)
97	Support Vector Machine classifier	(Steinberg <i>et al.</i> (2010)

CONCLUSION

In this study, a new approach to recognize a set of static hand gestures for the Human-Computer Interaction (HCI), based on hand segmentation using both wavelet

network and ANN has been proposed. All images in image database are divided into six hand gesture images classes (Zero, One, Two, Three, Four and Five). We have experimented with 60 hand gesture images and achieved higher average precision. The best classification rate of 97% was obtained the testing set.

REFERENCES

Al-Bashish, D., M. Braik and S. Bani-Ahmad, 2011. Detection and classification of leaf diseases using K-means-based segmentation and neural-networks-based classification. *Inform. Technol. J.*, 10: 267-275.

Bash, A.M., 2011. Predict dynamic response of suspension arm based on artificial neural network technique. *J. Applied Sci.*, 11: 988-995.

Chen, F.S., C.M. Fu and C.L. Huang, 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image Vision Comput.*, 21: 745-758.

Dardas, N.H. and E.M. Petriu, 2011. Hand gesture detection and recognition using principal component analysis. *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, September 19-21, 2011, Tianjin, pp: 1-6.

De Castro Fernandez, R.M. and H.N.D. Rojas, 2002. An Overview of Wavelet Transforms Application in Power Systems. *14th PSCC*, Sevilla, pp : 1-8.

Dong, G., Y. Yan and M. Xie, 1998. Vision-based hand gesture recognition for human-vehicle interaction. *Citeseer*, 1: 151-155.

Fang, Y., K. Wang, J. Cheng and H. Lu, 2007. A real-time hand gesture recognition method. *Proceedings of the IEEE International Conference on Multimedia and Expo*, July 2-5, 2007, Beijing, China, pp: 995-998.

Garg, P., N. Aggarwal and S. Sofat, 2009. Vision based hand gesture recognition. *World Acad. Sci., Eng. Technol.*, 49: 972-977.

Goswami, J.C. and A.K. Chan, 2011. *Fundamentals of Wavelets: Theory, Algorithms and Applications*. 2nd Edn., John Wiley and Sons, New York, ISBN: 9780470484135, Pages: 359.

Gu, Y., Y. Li, J. Xu and Y. Liu, 2011. Novel model based on wavelet transform and GA-fuzzy neural network applied to short time traffic flow prediction. *Inform. Technol. J.*, 10: 2105-2111.

Hasan, A.M., 2004. Algorithms for detection and identification of noise in images and denoising using wavelet network. In: *Control & Systems Engineering* vol. Ms.c IRAQ: University of Technology.

- Hasanuzzaman, M., V. Ampornaramveth, Z. Tao, M.A. Bhuiyan, Y. Shirai and H. Ueno, 2004. Real-time vision-based gesture recognition for human robot interaction. Proceedings of the IEEE International Conference on Robotics and Biomimetics, Aug. 22-26, Shenyang China, pp: 413-418.
- Khanale, P. and S. Chitnis, 2011. Handwritten Devanagiri character recognition using artificial neural network. *J. Artif. Intell.*, 4: 55-62.
- Kim, H. and D.W. Fellner, 2004. Interaction with hand gesture for a back-projection wall. Proceedings of the Computer Graphics, July, 06, 2004, Institute of Computer Graphics, Technology University, Braunschweig, pp: 395-402.
- Kurtenbach, G. and E.A. Hulteen, 1990. Gestures in Human-Computer Communication. In: *The Art and Science of Interface Design*, Laurel, B. (Ed.). Reading, Mass: Addison-Wesley Publishing Co., Wasley, pp: 309-317.
- Lee, D.H. and K.S. Hong, 2010. A hand gesture recognition system based on difference image entropy. Proceedings of the 6th International Conference on Advanced Information Management and Service, November 30-December 2, 2010, Seoul, pp: 410-413.
- Lekutai, G., 1997. Adaptive self-tuning neuro wavelet network controllers. Electrical Engineering Department. Ph.D Thesis, Blacksburg, Virginia: Virginia Polytechnic Institute and State University.
- Malik, S. and J. Laszlo, 2004. Visual touchpad: A two-handed gestural input device. Proceedings of the 6th International Conference on Multimodal Interfaces: State College, October 13-15, 2004, Pennsylvania, USA.
- Malik, S., A. Ranjan and R. Balakrishnan, 2005. Interacting with Large Displays from a Distance with Vision-Tracked Multi-Finger Gestural Input. University of Toronto, Toronto, pp: 43-52.
- Melam, K. and C. Amar, 2010. A wavelet networks approach for image watermarking. *Int. J. Comput. Intell. Inf. Secur.*, 1: 34-43.
- Murthy, G. and R. Jadon, 2010. Hand gesture recognition using neural networks. Proceedings of the IEEE 2nd International Conference on Advance Computing Conference, February 19-20, 2010, Patiala, pp: 134-138.
- Muzhou, H. and H. Xuli, 2010. The multidimensional function approximation based on constructive wavelet RBF neural network. *Applied Soft Comput.*, 11: 2173-2177.
- Palaniappan, R., M. Paulraj, S. Yaacob and M.S.B.Z. Azalan, 2010. A simple sign language recognition system using affine moment blur invariant features. Proceedings of the International Postgraduate Conference on Engineering, October 16-17, 2010, Perlis, Malaysia, pp: 1-5.
- Sharma, A. and S. Agarwal, 2012. Temperature prediction using wavelet neural network. *Res. J. Inform. Technol.*, 4: 22-30.
- Steinberg, I., T. London, D. Di Castro, 2010. Hand gesture recognition in images and video. CCIT Report No. 763, EE Publication No. 1720. http://webee.technion.ac.il/people/dot/Steinberg_London_DiCastro_Tech_Report.pdf
- Taqi, A.Y. and H.A. Jalab, 2010. Increasing the reliability of skin detectors. *Sci. Res. Essays*, 5: 2480-2490.
- Wu, M. and R. Balakrishnan, 2003. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, November 2-5, 2003, Vancouver, Canada, pp: 193-202.
- Yedjour, D., H. Yedjour and A. Benyettou, 2011. Explaining results of artificial neural networks. *J. Appl. Sci.*, 11: 2855-2860.