Mahmuod H. AL-Muifraje

Electrical and Electronica Eng. Dept, University of Technology, Baghdad, Iraq.
drmahmood6@gmail.com

Thamir R. Saeed

Electrical and Electronica Eng. Dept, University of Technology, Baghdad, Iraq.
thamir_rashed@yahoo.com

Mosab A. Hassan

Electrical and Electronica Eng. Dept, University of Technology, Baghdad, Iraq.
mosabali53@yahoo.com

Received on: 20/10/2016
Accepted on: 20/07/2017

Design and Analysis of American Sign Language Classifier Based on n-tuple Technique

Abstract- *This work consists of three stages: the first stage is the collection of data (images) for 26 letters alphabet, which is 1040 images, 40 for each class (letter). The second stage is the Preprocessing (segmentation, filtering, crop, convert to a binary image and feature extraction) and the last stage is the classy one. Two filters have been used median and wiener2 to treatment two types of noise; the salt and pepper and Gaussian noise respectively. The classification has been satisfied by using an n-tuple classifier algorithm is 93.8462%, while it is 93.2692% with salt and pepper noise. In this context, the recognition is 94.2308% when the median filter is used for salt & Pepper noise, while, it is 93.8462% when to use a Wiener2 filter with Gaussian noise. The optimal tuple size is 4, and suitable training pattern is 40 per each class.*

Keywords- *ASL, ASL recognition, n-tuple, WNN.*

How to cite this article: M.H. AL-Muifraje, Th.R. Saeed, M.A. Hassan , "Design and Analysis of American Sign Language Classifier Based on n-tuple Technique," *Engineering and Technology Journal*, Vol. 36, Part A, No. 2, pp. ١٧٢-١٧٨, 2018.

1. Introduction

Hand gesture technique is a method of communication between people and hearing impaired people (deaf). In this context, it is useful for communicating with them fluently. Finding an experienced and qualified interpreters every time is a tough task. Moreover, people who are not deaf, never try to learn the sign language for interacting with the deaf people, then, deaf people will be isolated. But if the computer can be programmed in such a way that it can translate sign language into text format, the difference between the ordinary people and the deaf community can be minimized [1]. Now, an overview of the other authors works done in this area, where, the first proposed gesture recognition between human and computer in the middle of seventies [2]. While the recognized of static hand gestures against complex background has been done by support Vector recognition system in [3]. [4] was using a McConnel's idea of orientation histograms for recognizing static hand gestures of ASL. [5] has been used an Adaptive Neuro-Fuzzy Inference system (ANFIS) models, for the recognition of a given gesture. [6] Was developed a recognition system, which depends on shape analysis. This system uses modified Fourier descriptors for the recognition of hand shapes. In [7], the hand gestures recognition which is based on an innovative self growing and self-organized neural gas (SONG) network has been proposed. The features extraction from hand gesture image and then combine them by using an induction graph has been presented by [8]. [9] was presented an ASL by using artificial neural networks (ANNs) Back Propagation (PB), after a required preprocessing the height, centroid, area and distance of the centroid from the origin (top-left corner) of the image are used as features. [10] was used hand segmentation by a depth disparity feature and, a hierarchical mode-seeking method, then, Random Forest (RF) classifier is built to recognize ASL signs using the joint angles. [11] has been used inertial sensors as input devices for the gesture sensing or hand motion sensing. In this paper, ASL has been presented, where the present work based on the three stages: the data collection (images), and Preprocessing, which is involved; segmentation, filtering, crop, convert to a binary image and feature extraction. While classification by using n-tuple classifier has been represented the last stage. In addition, median and wiener2 filters have been used as denoising filters.

2. Proposed System

The proposed approach consists of three stages, which include data (image) acquisition, pre-processing and classification. It is often so, the general architecture of the hand language recognition system is shown in Figure 1.

DOI: https://doi.org/10.30684/etj.36.2A.8
2412-0758/University of Technology-Iraq, Baghdad, Iraq

Figure 1: Architecture of the proposed system

I. Data (Image) Acquisition

All the movements in the ASL database are static in nature except the two alphabets 'J and 'Z' [12] as shown in Figure 2, where, these gestures are dynamic which involves the movement of fingers. In this work, the image acquisition has been taken in real conditions to find the solution for problems associated with getting the desired images. The image database contains 1040 images for 26 classes of ASL alphabets, each class containing 40 images (pattern) with static consideration. These images were captured to three different persons and in different lighting conditions. The type of these images is (JPEG) format with size 4320×3240 and 2160×3240 pixels, samples of these images are shown in Figure 3.

II. Hand Image Pre-processing Stage

This stage consists of six processes for extraction the features to make the classifier work. Figure 4 shows these processes.

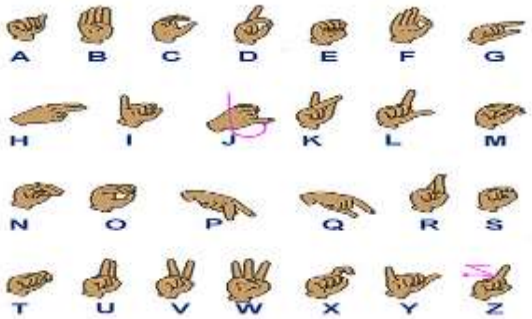


Figure 2: Hand language (American Sign Language Alphabet)

Figure 3: Representation of some ASL alphabet

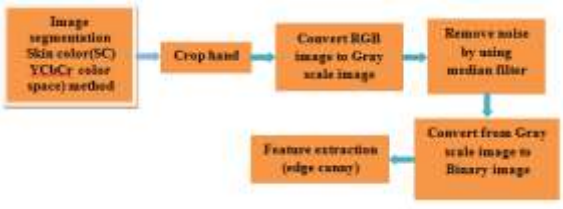


Figure 4: Flowchart of pre-proposed stage

III. Image Segmentation

The facial and hand colors features (YC_bC_r method) has been used to isolate the hand gesture from the background, which is called threshold segmentation algorithm. The luminance information can be represented by one component, Y, while, two components have been represented the color information, C_b and C_r. Where, C_b and C_r represent the blue and red color components. In this context, the transformation from RGB to YC_bC_r CS can be done by using the equation (1) as shown below [13]:

$$\left.\begin{aligned} Y &= 0.299\,R + 0.587\,G + 0.114\,B \\ C_b &= 0.169\,R - 0.332\,G + 0.5\,B \\ C_r &= 0.5\,R - 0.419\,G - 0.081\,B \end{aligned}\right\} \quad (1)$$

While Chai and Nagan algorithm has been used for skin representation by finding the value ranges of C_b and C_r as[12]: $77 \leq C_b \leq 127$ & $133 \leq C_r \leq 173$, Where, these values for Caucasian people skin (white human skin).

The proposed algorithm of skin color classification, as shown in Figure (5), has five stages these are:-

1- Transform the RGB image to YC_bC_r CS, as shown Figure (6a,b).

2- Consider the values of C_b and C_r as; ($77 \leq C_b \leq 127$ & $133 \leq C_r \leq 173$).

3- The Skin Pixel Quantifier counts the pixels number with human skin.

4- The classifying decision on the images depends on the percentage of skin content, if it is more than 50%, it is considered as an objection, as in equation (2) below, while if it is less than 50%, then the pixel is not considered as a hand object:

Skin percentage%= (skin color pixels/total image pixels) ×100. (2)

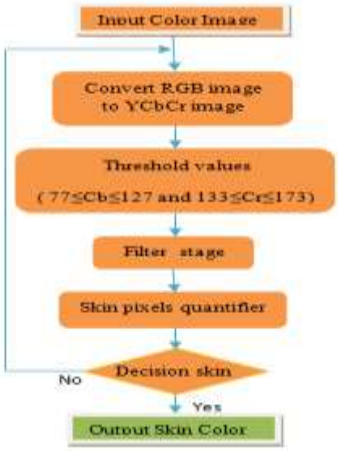


Figure 5: The proposed segmentation skin color system

Figure 6: Segmentation of Image (skin color)

After segmentation process, the noise has appeared in the form of points in the image due to different lighting. Then, to obtain the best image (low noise), a median filter has been used, where, it is a nonlinear operation often utilized in image processing for "salt and pepper" noise reduction, and it is more effective than convolution filter. It takes a slice [10 10] or [20 20] pixels of an image as shown in Figure 7.

IV. Hand Cropping

The proposed algorithm to crop the hand is based on the reading of images columns and rows, where, if the column value is zero, continue to read the next column and so on to reach the column value greater than zero. Where this reading has indicated arrival to hand, then measure hand dimension and crop the rest of the image as shown in figure (8(a,b)) Secondly, read the rows from top to bottom image, with the same procedure as shown in figure (8(c,d)). It follows that the required images size has been reduced to [110 × 72].

Figure 7: Filtering stage; a) image with noise, b) without noise

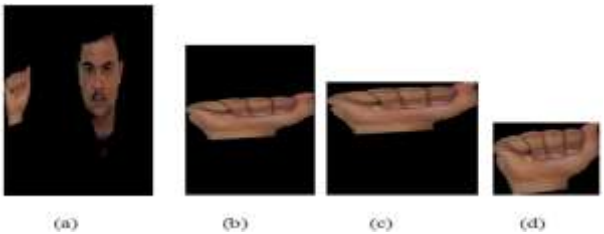


Figure 8: Hand cropping stages

Then convert RGB to the grayscale image by using Luminosity method, where, the method flowchart shown in Figure 9. Figure 10 shows the operation of this method on the target image.

To realizing the feature extraction, the image must be in binary form. The conversion process from grayscale to a binary image form is done by replacing all pixels in the grayscale image with luminance greater than a certain level (threshold) with the value 1 (white) and replaces all other pixels with 0 (black). Specify level in the range [0,1]; Otsu method has been used for obtaining the threshold value. From testing the image level, the threshold value of that image is found almost equal to the value of (0.4). In this context, this value does not give the required binary features of the hand image, so the optimal threshold value for this work is (0.3) because it gave more features to the edges of the hand show Figure (13(b,c)). Then, a binary image has been created, where, turning all pixels below the threshold to zero and the remaining pixels to one. Figure 11 shows the flow chart of this operation, while, Figure 12 shows converting (grayscale to binary) images.

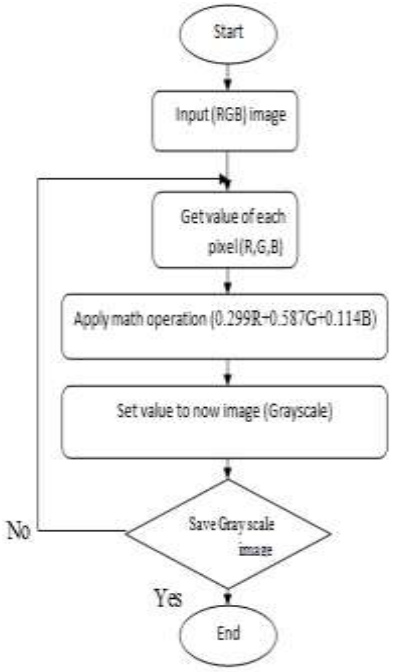


Figure 9: Flow chart shows the step of converting RGB image to grayscale

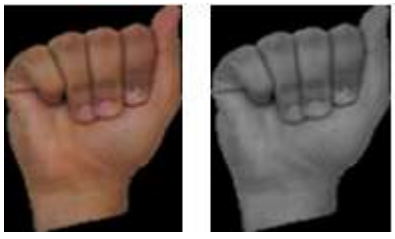


Figure 10: Convert RGB to Grayscale image

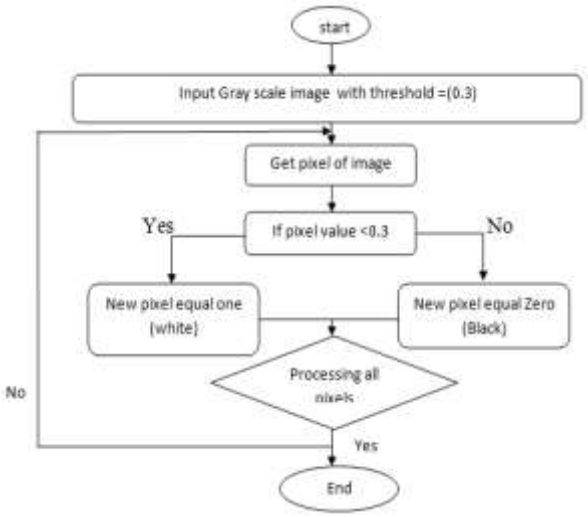
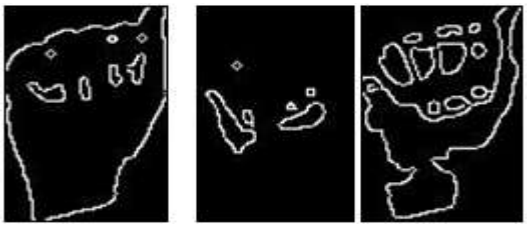


Figure 11: Flow chart shows the step of converting grayscale image binary

(a)Grayscale image (b) Binary image

Figure 12: Converting grayscale to binary image



(a) Thr.=0.3 (b) Thr.=0.4 (c) Thr.=0.5

Figure 13: Several threshold (Thr.) values

3. Features Extraction

The Static hand gestures, features represent the fingers direction; fingertips and hands contours, which can be extracted. In this work, the hand contour is a well-chosen feature to describe the hand gestures shape and select the smooth edge to this task, as shown in Figure 14.

Figure 14: Canny edge images

4. Classification

The classification is the last stage of the proposed system. Where the operations were performed on the hand images are made it ready to classification. The dimensions of each image are [110 72], in binary form. Then, these images will be classified by n-tuple classifier.

I. N-tuple Classification

This method consists of two stages these are:

A-Training stage

Figure 15 shows the training and saving the image data in the discriminator RAMs. The training stage starts, when each discriminator is trained on all the patterns of their class, by setting the bits access by n-tuple in the discriminator, related to that class.

Four n-tuple size '4, 6, 8 and 10' have been used to recognize the same test patterns for different training set sizes. It was found that n-tuple size "4" gives optimum recognition performance for the used pattern classes. This size, in fact, offers a good and optimal compromise between recognition capability and RAM size, which increases rapidly as n-tuple size increases. Thus, n-tuple size of "4" will be employed here. The mapping used for the construction of the n-tuple is random mapping (i.e. the n-tuple elements are chosen randomly from the input pattern pixel matrix), and it is more likely to detect general features.

B- Classification (testing) stage

The classification (testing) stage is an introduction to the patterns of various classes, which are considered for testing. All the discriminator will be introduced to the patterns, and the decision will be taken from the maximum response detector, which can detect the higher response between the twenty-six discriminations. Figure 16 shows classification stage.

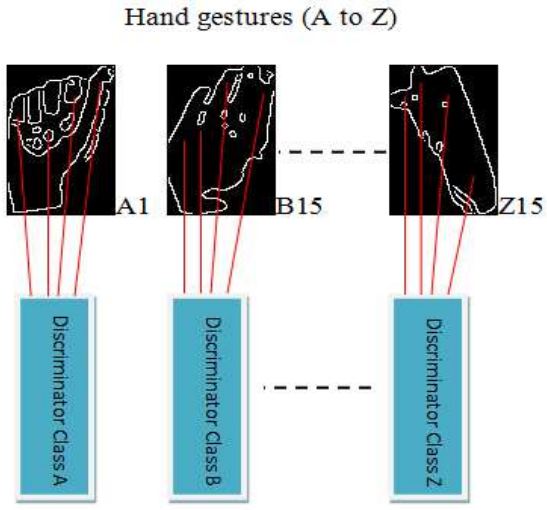


Figure 15: Training image in tuple system

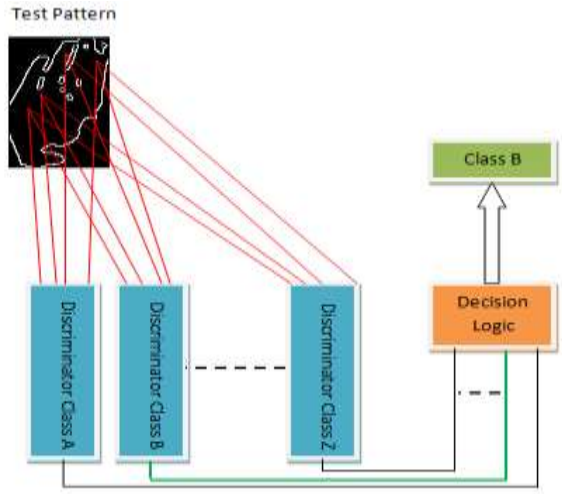


Figure (16) Classification of image in tuple system

5. Results

The goal of this work is to recognize the static hand gestures, which represent the ASL by using n-tuple method. In effect, there are four important factors impact on this recognition; these are; the number of a tuple, number of pattern per class, number of training images and the effect of the noise.

The various tests for tuple sizes "4, 6, 8, 10" have been resulted in a suitable size is "4". Where it is appropriate for feature extraction and system complexity and then is suitable for classification. In this context, the dimensions of the input pattern are [110 72] for all characters of American Sign Language. While the number of images in two phases, training, and testing, is 1040 pattern (520 patterns for training and 520 for testing) in all classes. Therefore, the results have been indicated the better performance of the system, is 93.8462%, while, the error recognition is 6.1538%. Table 1 shows the results of different tuple sizes while it indicates the preference of tuple 4 for best results by using (1040 pattern for all classes).

Consequently, an experimented was made to test the performance of 4-tuple with a different amount of training image (5, 10, 15, 20, 30, 40). The result as shown in Figure 17 is illustrated that the performance is rising with increasing of the training images size.

Table 1: Classification using different n-tuple Size

Classification of particular n-tuple				
Tuple size	4	6	8	10
Correct%	93.8462	93.6538	93.6538	93.4615
Error%	6.1538	6.3462	6.3462	6.5385

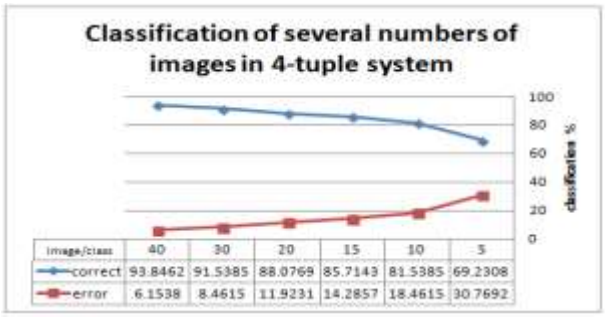


Figure 17: Results of several numbers of images using 4 –tuple

Now, the noise of salt & pepper has been added to find out the effect on the results and for an approach to the real environments. The effects of the adding noise are by noise density (d). Figure (18) illustrates the performance 4-tuple system (classification) with noise salt and pepper type with different densities.

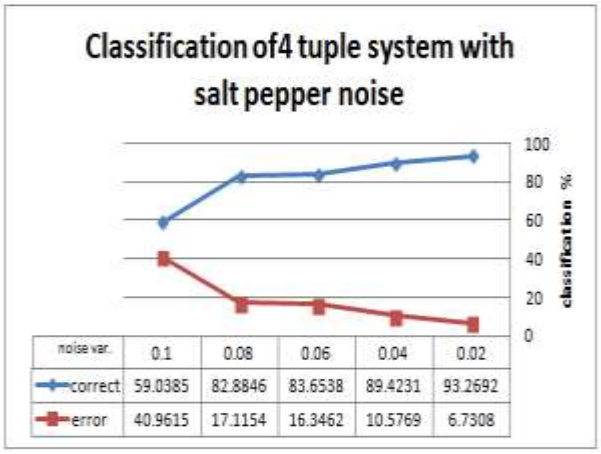


Figure 18: Classification of 4-tuple system with salt and pepper noise

Then, for reducing the effect of this noise, the median filter has been used. Table 2 shows the over-performance caused by using this filter, which has reduced the noise.

Another type of noise (Gaussian noise) has been added with mean=0 and several values of variance for experiment its effect on the 4-tuple system classifier performance. Figure 19 has shown the Collapse of the classifier performance. One interpretation of this influence is the use of inappropriate filter

The comparison of two filters; median and wiener 2, are shown in Table 3. The existing heuristics based classification appears the performance of filters with the two noise types (Gaussian and Salt and Pepper), and it is clear the advances of the wiener2 filter at the Gaussian noise. For putting the present work in a suitable place, the comparison with other work has been done as in Table 4. -year; R.M.-Recognition Method; N.O.C - Number of Class; NO.P -Number of Pattern; R.R-

Resonation Rate. From Table 4, the process of getting better is caused by using a proper feature extraction and classification algorithms.

Table 2: Effecting of used median filter to reduce salt and pepper noise

4-Tuple system median filter &noise salt and pepper		
d	Correct%	Error%
0.02	94.0385	5.9615
0.04	93.0769	6.9231
0.06	93.0769	6.9231
0.08	93.6538	6.3462
0.1	93.4615	6.5385

Table 3: The uses filters in tuple system

Tuple size is '4' (1040 images)			
Filter type	**Correct %**	**Error%**	**Correct without filters%**
Median only [3 3]	94.2308	5.7692	93.8462
Median filter [3 3] with salt and pepper noise (0.02)	94.0385	5.9615	93.2692
Wiener2 only [3 3]	92.88	7.12	93.8462
Wiener2 filter [3 3] with Gaussian filter(0.02)	91.7308	8.2692	67.5

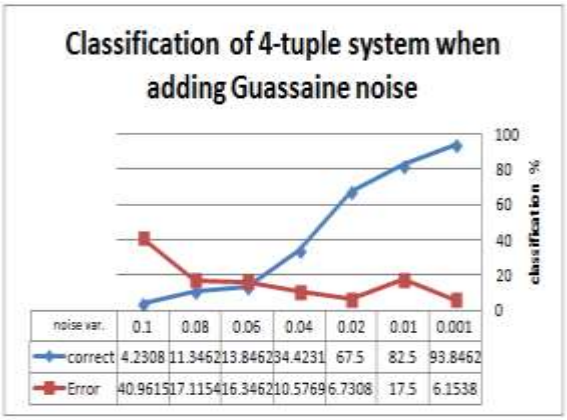


Figure 19: Classification of 4-tuple system when adding Gaussian noise

Table 4: Compared with previous works

Ref. No.	Y	R. M	N. O C.	NO. P	R. R
14	2011	Artificial neural network (ANN)	26	23	82.28%
15	2012	real-time hand gesture system	26	100	90.19%
16	2013	Support Vector Machine	10	5	92%
17	2015	a Random Forest (RF) classifier is built to recognize ASL signs using the joint angles	24	1	90%
11	2016	hand motion sensing and Dynamic time warping algorithm	10	1	87.5 %.
ourwork	2016	Present work used N-tuple method	26	40	94.23%

6. Conclusions

The aim of this work is to classify the static hand gestures, which represent the ASL by using n-tuple method and it is summarized in the following:

1. The optimal tuple size is 4, where the recognition of this size is 93.8462%, and when the median filter is used the recognition is improved to 94.2308%, while the Wiener filter is used the recognition decreases to 92.8846%.

2. The performance of the system is profoundly affected by the size of the training pattern. It is improved by increasing the training data. The recognition is 69.23% when using 65 training images, while the performance is improved to 93.846 when using 520 training images.

3. When adding Salt and pepper noise to images, the value (d = 0.02) chooses where the value of the recognition is 93.2692 and when using the median filter type ([3 3]), which has improved the discrimination to 94.0385. In addition, when adding the Gaussian noise to images the value (mean=0 and variance = 0.02) chooses where the value of the recognition is 67.5 and when using the filter wiener2 type ([3 3]), which has improved the discrimination to 91.7308.

4. From the results, it is clear that the median filter is the best for reducing the salt and pepper noise while wiener2 filter is suitable with Gaussian noise

References

[1] J. Singha and K. Das, "Indian sign language recognition using eigen value weighted Euclidean distance based classification technique," Assam Don Bosco University Guwahati, India, 2013.

[2] J.J. Stephan & S. Khudyer, "Gesture Recognition for Human-Computer Interaction (HCI)," Al-Nahrian University, Iraq, 2010.

[3] S. Naidoo, C. Omlin and M. Glaser, "Vision-Based Static Hand Gesture Recognition Using Support Vector Machines," 1998.

[4] K. Symeonidis, "Hand Gesture Recognition Using Neural Networks," School of Electronic and Electrical Engineering, thesis, August 23, 2000.

[5] O. Al-Jarrah and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," Jordan University of Science and Technology, Department of Computer and Internet Engineering, P.O. Box 3030, Irbid 22110, Jordan,2001.

[6] A. Licsar and T. Sziranyi, "Hand –gesture based film restoration," Hungarian Academy of Sciences, Hungarian, 2nd international conference Pattern recognition in information systems, PRIS'02, Alicante. JANUARY 2002.

[7] E. Stergiopoulou and N. Papamarkos, "A new technique for hand gesture recognition," Department of Electrical & Computer Engineering Democritus University of Thrace 67100 Xanthi, Greece, 2006.

[8] A.B. Jemaa, W. Mahdi, Y.B. Jemma and A. m. B. HMADOU, "A new approach for digit recognition based on hand gesture analysis," Higher Institute of Computer Science and Multimedia Sfax, Tunisia, 2009.

[9] M.A. Rahman, A.. Ambia, and M. Aktaruzzaman, "Recognition Static Hand Gestures of Alphabet in ASL," 2011.

[10] C. Dong, M. C. Leu and Z. Yin," American Sign Language Alphabet Recognition Using Microsoft Kinect," Missouri University of Science and Technology, 2015.

[11] P. Patil and S.S. Savkare, "Hand Motion Gestures For Mobile Communication Based On Inertial Sensors For Old Age People," Int. Journal of Engineering Research and Application, Vol. 6, Issue 9, Part -2, pp.50-54, September 2016.

[12] S. Nagarajan & T.S. Subashini, "Static Hand Gesture Recognition for Sign Language Alphabets using Edge Oriented Histogram and Multi Class SVM," Annamalai University, Annamalai nagar, Tamil Nadu, India, November 2013.

[13] B. Wang, X. Chang and C. Liu1, "Skin Detection and Segmentation of Human Face in Color Images," 1Hebei University of Technology, Tianjin 300130, China, 2011.

[14] K. K. Singh and A. Singh," A Study of Image Segmentation Algorithms for Different Types of Images," India, September 2010.

[15] R. Kandwal et al.," Review: Existing Image Segmentation Techniques," CSE Department, Dehradun Institute of Technology, India, April 2014.

[16] K. M.Khaki, "Weightless Neural Networks For Face and Pattern Recognition: An Evaluation Using Open-Source Databases"،," Ph.D. Thesis, Brunel University, London, August 2013.

[17] J. R. Armstrong," Boolean Weightless Neural Network Architectures, "University of Central Lancashire, April 2011.

Author(s) biography

Mahmuod Hamza AI-Muifraje received the B.Sc. in Electronic Engineering from University of Technology, Baghdad-Iraq in 1975. He received the M.Sc. and Ph.D. from Brunel university-UK in Electrical Engineering and electronics in 1979 and 1983 respectively. Currently, he is Assistant professor within the Department of Electrical Engineering at the University of Technology, Baghdad -Iraq. His main research interests include pattern recognition system, weightless neural networks, image processing and wireless sensor networks. drmahmood6@gmail.com

Thamir Saeed was born in Baghdad, Iraq, in February 1965. He received the BSc and MSc degrees both from Military Engineering College in Baghdad in 1987 and 1994, respectively. He was awarded his PhD degree from Al-Rashed College of Engineering and Science, Baghdad, in 2003. From 1994 to 2003, he worked with Military Engineering College in Baghdad as a member of teaching staff. From 2003 until now, he worked with the University of Technology, Iraq, as a member of teaching staff. Currently, he is an Assistant Professor of Electrical Engineering at University of Technology. His major interests are in Radar signal processing, digital circuit design for digital signal processing (DSP) based on FPGA, Image Processing, Pattern Recognition and Sensor network.
E-mail: 50257@uotechnology.edu.iq and thamir_rashed@yahoo.com

Mosab A. Hassan was born in Daila, Iraq, in February 1982. He received the BSc and MSc degrees both from UOT in Baghdad in 2005 and 2016, respectively. From 2016 till now, he worked with the University of Technology, Iraq, in reser h groups. His major interests are in Image processing and Pattern recognition. E-mail: mosabali53@yahoo.com