English Alphabets Recognition using Hand Gesture

Hanaa M. Ahmed, Ph.D.(Asst.Prof.)* **Yasser D. Hussein***

Abstract: In Human-Computer Interaction (HCI) area, the sign language is needed for individuals with special needs (dumb) for the purpose of communicating with individuals with no special needs and also for communicating with each other. Thus, the sign language is determined via using the systems of hand recognition. The initial stage in utilizing manual signal recognition is deleting the background in addition to a part (arm/hand) from other body parts just as much as feasible. Concerning the presented study, we are offering a novel approach for representing the alphabet of English language as 6 pattern gestures. The hash pattern has been applied through utilizing histogram equation for gray-scale image. The process of classification has been implemented via the Euclidean distance function as well as Fisher Linear Discrimination Analysis (FLDA). Difficulties including recognizing similar gestures appear to be managed quite efficiently with the proposed technology, as the background is fixed white. The recognition rate has been 82.86 since the lighting differs between different places, in addition to the differences in resolution of the camera, that is highly important in characterizing features.

Keywords: Hand gestures; Sign language; Fisher linear discriminant analysis; Human Computer Interaction; Euclidean Distance**.**

* Computer Science Department, University of Technology, Baghdad, Iraq

1. Introduction

Concerning the HCI field, these systems use the human hand as a natural input that afterwards turns out to be a major motivation for all studies carried out in the subject related to recognition by using the hand. The technology of Virtual Reality (VR) which helps individuals to have a feeling of involvement in computer world, has become a topic of public study for a lot of years until this time is considered as an example of high importance regarding using hands for the purpose of having interaction with computers via the sign language. It provides a user-friendly and appealing replacement to interface devices like joysticks, mouse, keyboards, and so on. [1], [2].

Vision based on hand gesture systems either carry out visual analysis of the shape of the hand, movement/position as a first type. The second hand-based gesture system type which is referred to as the structure which is based on the glove, however, is more optimal than a particular viewpoint but needs preparing tedious approaches and bounds the user's comfort, simplicity and naturalness of work. [3]. In HCI, for the sake of communicating with the devices, the fields of applications are identifying each of gesture and posture. A gesture can be characterized as a spatio-temporal pattern that can be static, dynamic or both. In addition to that, postures are static hand morphs and the movements of the hand are referred to as gestures [4].

In [5] the past, group of approaches and algorithms have been proposed by some researchers. In hand gesture recognition have been suggested by the researchers for controlling home appliances which could be divided into 2 methods: one of the methods is vision-based approach while the other is inertial sensor-based methods. In [6], the researchers offered extension to Kinect SDK on the basis of a contour analysis to estimate the hand position. The algorithms that have been presented by the researchers are then applied for providing the development of gesture library which could be utilized later. In [7], the researchers have suggested a model based approach to track the motion of the hand in space where the algorithm of Dynamic Time Warping (DTW) is applied for time-alignment as well as for normalizing spatiotemporal variations which exist amongst the samples that belong to one class of gestures. In [8] the researchers suggested a sufficient Machine Learning (ML) approach for analyzing multi-media content that addressed gesture event detection and recognition. The ML approach they have applied is implemented according to efficiently-studied approaches like applications to Smart TV virtual keyboard, Combination of Global and Local Representations, Linear Shape Model and Procuresses Analysis. In[9] they have created robust marker-less system of hand gesture recognition that might effectively track dynamic as well as static hand gestures. The identified gesture will be translated to action in their system like opening applications such as PowerPoint and VideoLAN Client (VLC) as well as opening websites.

In the presented study, a set of methods were reported for recognizing six hand gestures which are later utilized for representing all isolated English alphabets with adding space and out. An enhanced model of skin color segmenting has been utilized on the basis of the graph equation and gray-scale image. Tasks like the recognition of similar gestures in a variety of backgrounds and illumination conditions are taken. The classifier is implemented with the use of LDA, and then utilized to identify isolated English alphabet with the use of six lists.

2. Preliminaries

2.1 Image Enhancement

Optimizing the quality of the image through converting colored image to comprehensive gray-scale image for preserving the photo's highlights, including image structure, sharpness, shadow and differences. Through applying equation 1, the RGB color will be converted to gray-scale[10]:

$$\mathrm{Gray}_{\mathrm{Scale}} = (0.587 \times \mathrm{G}) + (0.114 \times \mathrm{B}) + (0.299 \times \mathrm{R}) \tag{1}$$

2.2 Equalization of Histogram

The method of contrast improvement is utilized according to traditional graph equation algorithm. Which is considered comparable to the histogram's equalization method which controls the outcome related to the graphing method in a way that it eventually enhances the image[11].

The histogram equalization equations are below:

$$pdf(x) = p(r_m) = \frac{total\ pixel\ with\ intensity\ rk}{total\ in\ image\ x} \tag{2}$$

The Cumulative Density Function (CDF) might be obtained from the above-described Probability Density Function (PDF) in the following representation:

$$cdf(x) = \sum_{m=0}^{L-1} p(r_m) \tag{3}$$

Where p(r_m) can be defined as the probability for the pixel of density r_m . p(s_m) must undergo multiplication by L-1, and after that, rounded to the nearest integer to obtain the pixel's value.

The general equation of histogram equalization for obtaining the value of the new pixel is depicted below:

$$S_m = (L-1) * \sum_{m=0}^{L-1} p(r_m) \tag{4}$$

2.3 Hand Segmentation

This approach is used to separate the hand of the user from the image's background. Thresholding is applied for achieving this goal. Depending on a threshold value, the thresholding will be implemented on the input image. The pixel will be set to one when its intensity is more than the threshold value, while the pixel will be set to zero when its intensity is less that the threshold value. The accuracy of hand detection might be reduced depending on the hand color of the user even in constant lighting environments throughout each system usage[12].

2.4 Cropping and Resizing the Image

Cropping is done following to the histogram equalization process. The cropping step is implemented for developing frame by removing the image's outside components. The resizing process is going to alter the input image's size to (70 x 70) pixels in dimension which is going to be applied for feature extraction. **Resizing process is considered quite different from the cropping process.** It implicates making an image "weigh" less, efficiently reducing the image's real size (MB), it is necessary to reshape the training phase's data for all frame's vectors to be of an equal size to equation (5) [13].

$$i_{\_resized}(k,l) = \sum_{y=1}^{k} \sum_{x=1}^{l} i\_grayscale(x,y) \tag{5}$$

Where; the Height and the Width related to the source frame are (y=240, x=320)

The Height and the Width related to the target frame are (k= l= 70)

2.5 Linear Discriminative Analysis (LDA)

Linear Discriminative Analysis is used to search for the vectors in the underlying space which optimally differentiates amongst classes. The major aim of this method is finding linear transform in a way that the feature clusters will be mainly separable following the transform that could be done via scatter matrix analysis. The objective of this method is to minimize within-class scatter matrix measure, and at the same time maximize between-class scatter matrix measure . There are two different approaches that will be used in the next two sub-sections for calculating the LDA space[14].

1- **Class-dependent method**: In this approach, single separate lower dimensional space will be determined for each one of the classes as follows, Wj $= \mathrm{Sw}^{-1}\mathrm{SB}$, as Wj represent transform matrix for j^{th} class. Therefore, eigenvectors and eigen values are going to be determined for each transformation matrix in a separate manner. Thus, each class's samples will be projected on their matching eigen vectors.

2- **Class-independent method:** In this approach, single lower dimensional space will be determined for the classes. Therefore, transform matrix will be determined for all of the classes, also the samples that are related to all classes will be projected on selected eigenvectors.

Where: the sample for all equation mean:
n_j = can be represented as the overall number of samples on the class
N= can be represented as the overall number of samples for all the classes
c: Class number, *mi*
μ: can be defined as the total mean value
Ni : mean of sample and corresponding classes size. Solve the popularized difficulty of the eigen value for the array $\mathrm{Sw}^{-1}\mathrm{SB}$, and after that, check the (eigen vector-eigen value) computation
W= $\mathrm{Sw}^{-1}\mathrm{SB}$
V: is the Eigen vector
λ: is the Eigen value
The first stage involve calculating the mean vectors for each one of the classes, m_j (j=1, 2, 3), the

Mean will be written in a form of a matrix [1×M]. Represented by the following equation[15]:

$$m_j = \frac{1}{n_j} \sum_{x_i \in \omega_j}^{\mathrm{n}} x_i \tag{6}$$

The equations below represent calculating the total average for all data

$$m = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} x_i = \sum_{\mathrm{j}=1}^{\mathrm{c}} \frac{n_j}{\mathrm{N}} \mu_j \tag{7}$$

The following phase is computing scatter matrix contains between-class scatter matrix (11) and within–class scatter matrix (9), (10).

$$S_{w_j} = d_j^T * d_j \tag{8}$$

The values that are related to within-class matrix for every class and the total within-class matrix are represented by:

$$S_w = \sum_{\mathrm{j}}^{\mathrm{c}} S_{w_j} \tag{9}$$

The between-class scatter matrix will be calculated as follows :

$$S_B = \sum_{\mathrm{j}}^{\mathrm{c}} \mathrm{Ni}(\mathrm{mi} - \mathrm{m})(\mathrm{mi} - \mathrm{m})^{\mathrm{T}} \tag{10}$$

Solve the popularized eigen value difficulty for array $\mathrm{Sw}^{-1}\mathrm{SB}$, and after that, verify (eigen vector-eigen value) calculation:

$$W\ V = \lambda\ V \tag{11}$$

The final phase, the initial data will be projected on lower dimensional space in the following way, where $y_j(n_j \times 1)$ represent data after projecting the i^th class

$$Y_j = \omega_j V_k \tag{12}$$

The equations below represent calculating the Recognition rate.

$$\mathrm{RR}\ (\%) = (\mathrm{C/T}) \times 100\% \tag{13}$$

3. Proposed Approach

All previous steps that represent in diagram that are explained as in the figure (1).

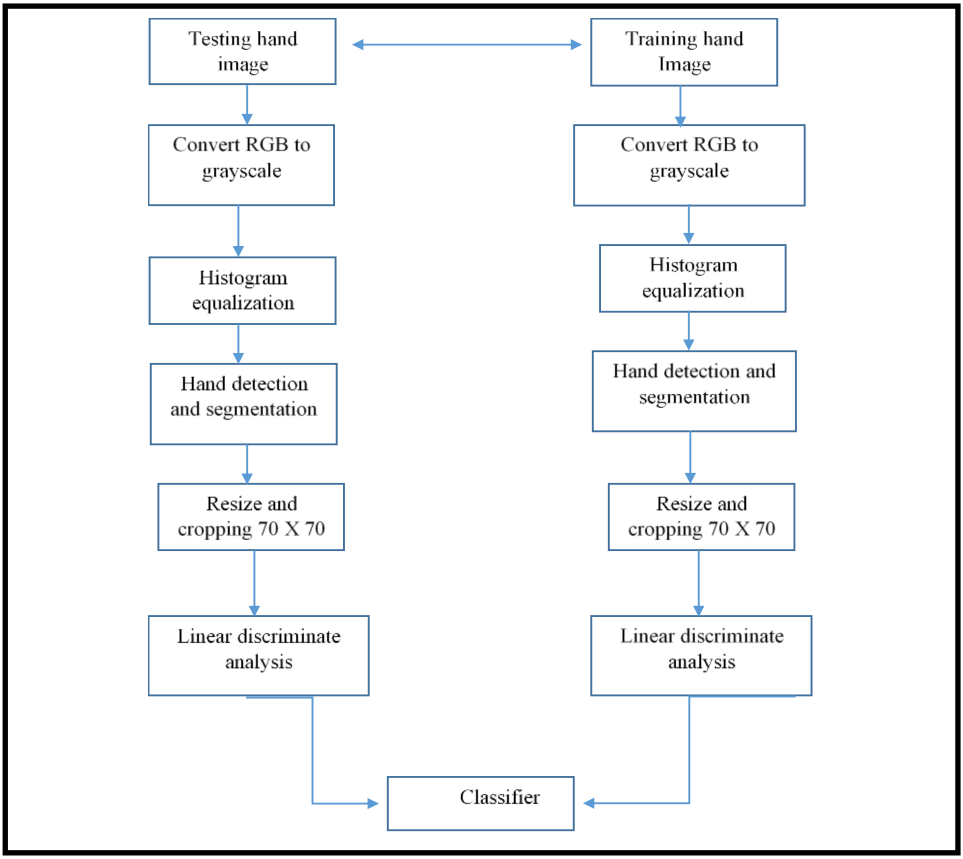


Figure 1. Hand gesture recognition system diagram

3.1 Convert RGB to Grayscale Image

In this section each resized data image has RGB shading space and it will be changed over into grayscale image. Grayscale is images preparing technique that take a three channel Red, Green, and Blue of the input hand gesture and changes it to a unique channel of gray value. The big picture of this process is addressed in vector structure as (Red, Green, and Blue). The RGB color is converted to grayscale using the previously entitled equation (1).

Algorithm (1): Converting Color frame into Grayscale
Input: Color frame //captured color frame
Output: Grayscale frame // converted color frame into grayscale frame
BEGIN **Step1:** Read color frame **Step2:** Convert the color frame into grayscale frame For x1 = 0 to x−1 the frame width For y1 =0 to y−1 the frame height Grayscaleframe(x1,y1)=0.2989*R(x1,y1)+0.5870*G(x1,y1)+0.1140*B(x1,y1); End for y1 End for x1 **Step3:** Return gray frame **End**

3.2 Histogram equalization of image

The aim of histogram equalization is evenly distributing the contrast of a certain image over the entire available dynamic area, which means between zero and one. In the approach of histogram equalization, it is the probability density function (PDF) being altered. It alters the PDF of a certain image to that of a uniform PDF which spreads out from the smallest values of a pixel value (i.e. zero in this case) to the maximum pixel value (L-1). Where L is gray level of image is equal 255. This may be accomplished quite easily in the case where the PDF is a continuous function. On the other hand, since a digital image is used, the the probability density function will be a discrete function. Supposing that an image x and let the dynamic range for the its intensity rk vary from zero (i.e. black) to L-1 (i.e. white). This the probability density function might be approximated with the use of the probability according to the histogram p(rk) based on the Previously entitled E.q(2).

From this the probability density function, then calculate the cumulative density function (CDF) according to Previously entitled E.q (3).

Where p (r_k) denotes the probability for pixel of density r_k . for the aim of getting the pixel value, p (s_k) must be multiplied by L-1 and after that rounded to the nearest integer.

That was a general equation for histogram equalization to obtain the new pixel values as follows on Previously entitled E.q(4).

The following shows applying the result of steps histogram equalization as in Figure (2):

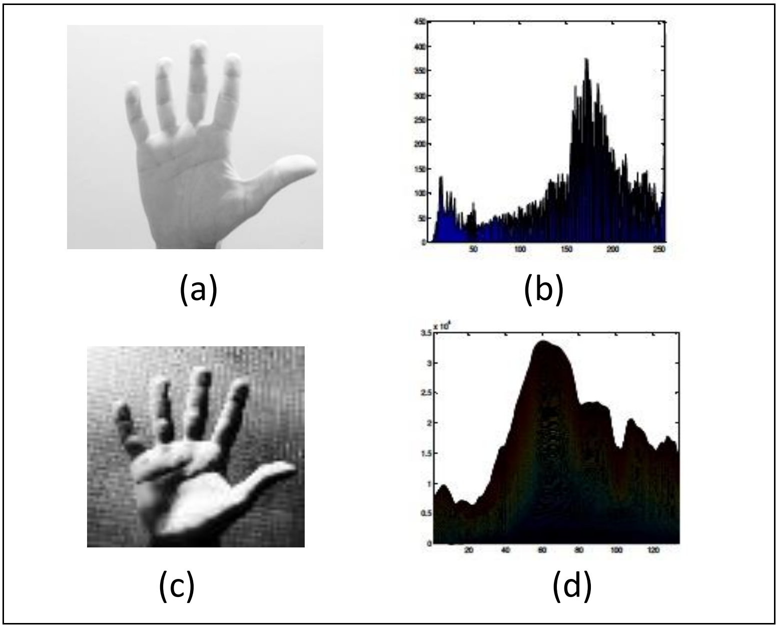
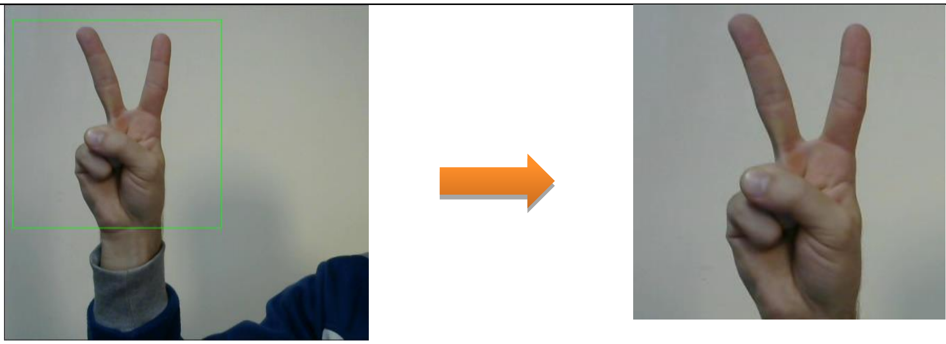


Figure 2. Histogram equalized where (a) hand image that is of a low contrast, (b) hand image Histogram, (c) Histogram equalized hand image low contrast (d).

3.3 Cropping and Resize Stage

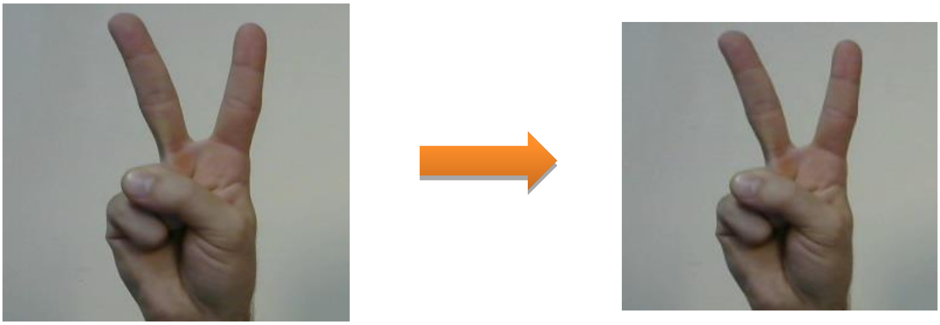
In this stage a box of red color frame is drown over the screen, and tries to position users hand inside the drown red frame. Then eliminates exterior components of an image to develop the frame. Cropping the hand image as shown as following in Figure (3):



a. Original image b. Cropped image

Figure 3. a, b Image Cropping

The resizing procedure will alter the input image size to (70 × 70) pixels in dimension which will be utilized to Convert RGB to Grayscale Image. For the following Previously Entitled E.q (2.5) as in Figure (4):



(a) Original Frame (b) Resizing Frame

Figure 4. a, b Image Resizing

3.4 FLDA Features Extraction method

The fundamental objective of the FLDA is to find a base of vectors that provides the optimal differentiation amongst classes, attempting to increase the differences amongst classes, decreasing the differences within classes. FLDA takes under consideration the different hand variables and determines and decides to which one of the groups the hand is most possibly a part of. The FLDA generates a group of feature vectors where variations of different hands are pointed out, whereas different hand instances because of lighting conditions, hand orientation and expression are de-emphasized.

The FLDA algorithm is applied to all hand images as in Algorithm (2). The results are used later as feature vector after it normalized into range (0-1).

The results result feature vector moves to training step/testing step as in following work FLDA by steps on this algorithm:

Algorithm (2): Work FLDA
Input: hand images dataset
Output: hand images trained
BEGIN **Step 1:** Acquire the training set of hand images and from its feature vector representation (hand gesture image). **Step 2:** Obtain the within class and between-class covariance matrices Sw and Sb according to equation (9) and equation (10). **Step 3:** Obtain the transformation matrix with higher reparability according to equation (12). **Step 4:** Classify the input hand image, based on the values of the Euclidean distance and threshold. **END**

4. Dataset of hand gesture

There are six lists (group split) to represent the alphabet of English language plus space and out (represented elements). There are five alphabet characters in each list, while the sixth list contains the space and Z as in the figure (5). Each list is represented by each gesture of the hand. Hand gesture image is used to represent any list element. Thus, for representing any element one (two) hand gesture must be used, which is based on list and the character which needs to be represented. As it is seen in Figure (5).

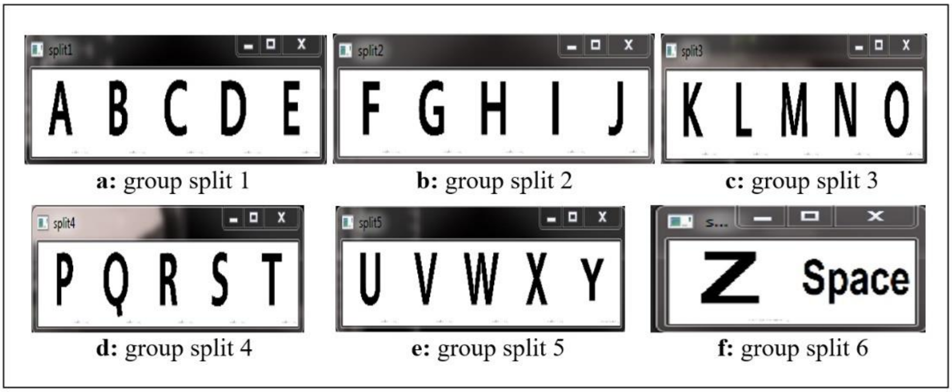


Figure 5. a to f six group splits

The hand gesture recognizer could be implemented through the use of pre-captured hand gesture images (frames) taken via camera of laptop. A red color square will be drawn over the screen of the computer to assist segmenting the hand image. The size regarding the pre-captured hand gesture images (prepared data-set) consist of sixty hand images, for six distinct samples of hand images, as seen in Figure. (6).

Reference1	Reference2	Reference3	Reference4	Reference5
				
				

Figure 6. Dataset of hand image

Images go to conversion from RGB to Grayscale, histogram equalization, cropping and resizing to (70×70), and after that, LDA can be used for feature extraction. Lastly, for implementing the classifier, the Euclidean distance will be applied for intra and inter class recognition. For the prepared data-sets (training phase), the resulting LDA classifier have the ability to determine the six different isolated hand gestures with rate of one hundred percent. Decision recognition (selecting list/choosing element) is based on input captured frame to the Euclidean distance and to the classifier.

5. Experimental Results

The LDA classifier is applied for identifying (resonation) isolated alphabets in runtime. For identifying the English alphabets, the general achieved recognition rate has reached (82.86%). In table (1), confusion matrices are depicted for the isolated alphabets recognition utilizing hand gestures with true and false recognition, and recognition rates. The formula specified in (13) is used to express the Recognition rate (RR) in %:

Where ‘T’ can be defined as the overall number regarding the tested videos sample, while ‘C’ can be defined as the correct number of spotting.

In this paper, ten sample videos have been tested for every one of the isolated gesture alphabets.

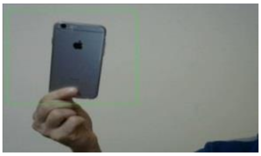
From Table (1), the output for overall RR for isolated gestures will be: (2320/28) × 100% = 82.86% , This percentage do not affect the accuracy of the proposed system, which is caused by the possibility of deleting the character if the identification is wrong. So the recognition ratio is good.

Table 1. Isolated gestures confusion matrix

Gestured characters	False	True	RR(%)
A	2	8	80%
B	1	9	90%
C	2	8	80%
D	3	7	70%
E	0	10	100%
F	3	7	70%
G	2	8	80%
H	3	7	70%
I	3	7	70%
J	1	9	90%
K	1	9	90%
L	3	7	70%
M	0	10	100%
N	2	8	80%
O	1	9	90%
P	3	7	70%
Q	2	8	80%
R	1	9	90%

S	1	8	80%
T	0	10	100%
U	3	7	70%
V	2	8	80%
W	1	9	90%
X	3	7	70%
Y	1	9	90%
Z	2	8	80%
SPACE	1	9	90%
out	0	10	100%

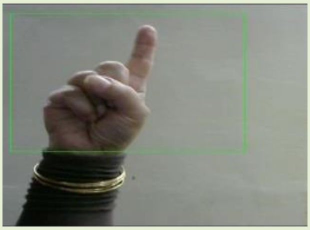
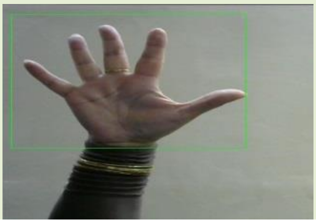
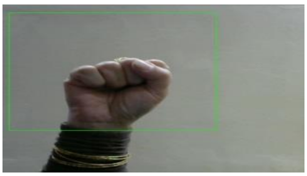
The test of the system in different and varied things was not accepted and not show any letter appear and identified as shown in figures(4):

Images	Result
	Not Recognize
	Not Recognize
	Not Recognize

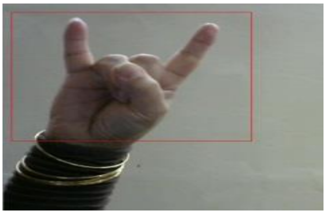
	Not Recognize

Figure 4. Testing different things

The proposed method accepts all hands no matter how skin color or shape of the hand whether a man or a women. Therefore, the public is considered authoritative for all. The following figure shows test images of the shape of a woman's hand as shown in the following figure (5):

Images	Result
	Recognize
	Recognize
	Recognize
	Recognize

Recognize



Recognize

Figure 5. Hand gesture for women (That Different hand from Dataset)

5.1 Comparison with previous works

The recognition rate achieved for isolated English is 81% using HMM. Furthermore, the recognition rate achieved for isolated English alphabets and The recognition rate achieved for isolated English is 81% using ANN is 85% for isolated and 84% for continues Considering the fact that, unlike HMM, CRF or DTW, gesture recognition systems are yet to use significant applications of neural networks, and as a result seldom approached with it, the recognition rate is satisfactory. So the percentage of this proposed system FLDA recognition rate is 82% for all isolated English alphabet . From this Table (2) the full Comparison with other methods.

Table 2. Comparison with other methods.

	Method	Isolated/continues English alphabet	English alphabet	Result %
1	**ANN**	**isolated**	**6**	**85%**
2	**ANN**	**continues**	**6**	**84%**
3	**HMM**	**isolated**	**All**	**81%**
4	**FLDA**	**isolated**	**All**	**82%**

6. Conclusions

In the presented study, a novel approach has been suggested for recognizing isolated alphabets of English language by applying six different hand gestures. LDA has been applied to assist the recognition procedure. The achieved recognition rate for the isolated gestures is 100%, which was because of utilizing small size data-sets (reference size (ten) for each of the six samples). The system proved to be efficient in handling issue of hand gesture recognition in various environments of lighting and backgrounds. The suggested system has the ability to recognize isolated alphabets with 82.86% accuracy. The rate of recognition which is related to space and out has been 100%, that will be very helpful to apply in English writing utilizing hand gestures, which will be considered as the Future work.

The contributions that are offered to the suggested approach are, first minimizing the samples of hand gestures to six, while the previous approaches utilized samples of size equal to the language alphabets. Second lists have been suggested for mimicking hand gesture based virtual keyboard for the proposed system with six keys. Third red square has been applied for the purpose of easing hand image segments from other parts of human body. Lastly, a method has been discovered for communication via texting between deaf individuals and non-deaf ones and state their emotions, and that is the most vital, so far.

References

[1] M.K. Bhuyan, P.K. Bora, and D. Ghosh, “Trajectory Guided Recognition of Hand Gestures having only Global Motions,” World Academy of Science, Engineering and Technology, vol. 21, pp. 753- 764, September 2008.

[2] Z. Yang, Y. Li, W. Chen, and Y. Zhang, “Dynamic Hand Gesture Recognition Using Hidden Markov Models,” International Conference on Computer Science and Education, pp. 360-365, July 2012.

[3] X. Zabulis†, H. Baltzakis†, A. Argyros, “Vision-based Hand Gesture Recognitionfor Human-Computer, InteractionComputer Science Department, University of CreteHeraklion, Crete, Grece.

[4] Pramod Kumar Pisharady and Martin Saerbeck, “Recent methods and databases in vision-based hand gesture recognition,” Institute of High Performance Computing, A*STAR #16-16 Connexis, 1 Fusionopolis Way, Singapore-138632

[5] Jiahui Wu1, Gang Pan1, “Gesture Recognition with a 3-D Accelerometer”, Department of Computer Science Zhejiang University, Hangzhou, 310027, China

[6] A. Lekova, D. Ryan, R. Davidrajuh, “Fingers and Gesture Recognition with Kinect v2 Sensor,” Print ISSN: 1312-2622; Online ISSN: 2367-5357 DOI: 10.1515/itc-2017-0009

[7] M.K. Bhuyan, P.K. Bora, and D. Ghosh, “Trajectory Guided Recognition of Hand Gestures having only Global Motions” International Journal of Computer Science 3;4

[8] Dequan Zheng , “Research of Machine Learning Method for Specific Information Recognition on the Internet “ , ICMI '02 Proceedings of the 4th IEEE International Conference on Multimodal InterfacesPages 229

[9] Aashni Haria , “Hand Gesture Recognition for Human Computer Interaction”, IEEE, International Conference on Signal Processing and Integrated Networks· February 2015

[10] Kanan C, Cottrell GW, “Color-to-Grayscale: Does the Method Matter in Image Recognition” , Eshel Ben-Jacob, Tel Aviv University, Israel

[11] Song, Ki Sun; Kang, Hee; Kang, Moon Gi, Jyothi S Nayaka* , “Hue-preserving and saturation-improved color histogram equalization algorithm.” , Science.gov is a gateway to U.S. government science information

[12] Senthilkumaran N1and Vaithegi S* , “IMAGE SEGMENTATION BY USING THRESHOLDINGTECHNIQUES FOR MEDICAL IMAGES.” , Department of Computer Science and Application, Gandhigram Rural Institute, Dindigul 2Deemed University, Gandhigram, Dindigul

[13] Weiming Dong; Xiaopeng Zhang; * , “Fast Multi-Operator Image Resizing and Evaluation.” , Journal of Computer Science and Technology 27(1):121-134

[14] Satoshi Niijima and Yasushi Okuno * , “Laplacian Linear Discriminant AnalysisApproach to Unsupervised Feature Selection.” , IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 6, NO. 4,

[15] Tom M. Mitchell* , “Reducing Data Dimension.” , Bishop, chapter 3.6, 8.6•Wall et al., 2003

تطبيق التواصل من خلال الرسائل المبني على اشارات اليد

أ. م. د. هناء محسن احمد* **ياسر ضياء حسين***

المستخلص: في مجال تفاعل الكمبيوتر مع البشر (HCI)، وهو اختصار للتواصل بين الإنسان والحاسوب، هناك حاجة إلى لغة الإشارة للأفراد ذوي الاحتياجات الخاصة (البكم) لغرض التواصل مع الأفراد الذين ليس لديهم احتياجات خاصة وكذلك للتواصل مع بعضهم البعض. وبالتالي ، يتم تحديد لغة الإشارة عن طريق استخدام أنظمة التعرف على اليد. المرحلة الأولية في تطبيق التعرف على الإشارات اليدوية هي حذف الخلفية بالإضافة إلى جزء (ذراع / يد) من أجزاء الجسم الأخرى بقدر ما يكون ذلك ممكنًا. فيما يتعلق بالدراسة المقدمة ، نحن نقدم نهجا جديدا لتمثيل الحروف الأبجدية للغة الإنجليزية على أنها ايماءات باستخدام ست أنماط. تم تطبيق نمط التجزئة للصورة ذات المقياس الرمادي. ومن ثم تم تحسين الصورة وزيادة نسبة التباين في الصورة من خلال معادلات ال histogram وكذالك تم تنفيذ عملية التصنيف عبر دالة تحليل التمييز الخطي في فيشر .(FLDA) ويبدو أن الصعوبات ، بما في ذلك التعرف على إيماءات مشابهة ، تدار بكفاءة تامة مع التكنولوجيا المقترحة ، حيث أن الخلفية ثابتة باللون الأبيض. أن سبب الحصول على معدل تمييزوصل الى 82.86 لأن الإضاءة قد تكون مختلفة من مكان لاخر ، بالإضافة إلى اختلاف دقة الكاميرا ، وهذا أمر مهم للغاية في تمييز الميزات.

الكلمات المفتاحية: إيماءات اليد، لغة الإشارة فيشر تحليل خطي التمييز، تفاعل الإنسان والحاسوب، المسافة الإقليدية.

* قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.