International Journal of Engineering & Technology, 7 (4.25) (2018) 291-297

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



An algorithm to Convert the Gestures of Numbers of Four languages into Voice

Shaker K .Ali[1], Hamdan Lateef Jaheel[2]

[1,2] *Computer Department, Computer sciences and mathematics college, university of Thi_Qar. Iraq*

Abstract

Gestures are one of the best ways of communication between dumbs people they can speak and other people they can speak using the expression of signs language. It is good to communicate between the dumb and other people by convert their signals into voices to be easy communicate with other people they can speak and hear voices. In this paper, a new algorithm proposed for recognizing hand gestures of numbers (0-10) to four languages (English, Arabic, Chinese and Persian languages) depending on dumbs signs and convert the signs into voices corresponding to signs numbers of each language. The proposed algorithm, firstly uses video for gesture of the dumb to three languages then converts these videos into frames (images) , secondly preprocessing step to removing the noises, resizing the images and increasing the contrast, the third step is extraction features step to calculating the distance of clustering algorithms such as Bayesian, C4.5, k-mean, k- medoid and artificial neural network. Eighteen features are calculated; eight features from Euclidean distance, eight features from slop, Area, and perimeter. The results in the training stage were; Bayesian gave 100% accuracy, C4.5 gave 100% accuracy, k-mean gave 100% accuracy k-medoid gave 100% accuracy and artificial neural network gave 95% accuracy. While in the testing stage classifiers are; Euclidian Distance, Modify Standardize Euclidian Distance and Correlation to calculating the difference between the features stored from training stage with new tested features and the results show that Euclidian Distance gave 100% accuracy, modified Standardize Euclidian Distance gave 100% accuracy and Correlation gave 100% accuracy. The database is created in our laboratory (six videos with 324 frames).

Keywords: *Gestures, Feature Extraction, Bayesian, C4.5, K-Mean, K-Medoid, ANN, Numbers, Clustering.*

1. Introduction

Communication is the way of people for expression Their ideas, thoughts, opinions or information between the people by writing, speaking, or signs. The dealing with people may be oral expression between ordinary people, while the dumbs people can't speak with ordinary people they can use signals to deal with the ordinary people [1].

Communication between people can be though the language to make them understand each other, but different languages may cause misunderstand the other people whom not native language such as Chinese language different from Arabic language, this problem may be solving by using translators for all languages, the dumb people have problem to communicate with normal people.

The gesture is a kind of nonverbal communication with a part of body such as hand or two hands signs for numbers, which used together with verbal communication. Same of people able to speaking and writing, but the others (dumb) could not only the can used gestures to deal with others and the gesture may change from person to another person, even the same person may change his gesture in some cases [2].

Dumb people use signs to express their ideas. Gesture language is different from country to another with its special vocabulary and grammarian. In fact, gesture language may be varied in the same country from place to another place as Languages spoken [3] so we used four languages (English, Arabic, Chinese and Persian languages). The gesture is the action of any part of hand or two hands motion [4].

There are many methods for recognizing the gesture; such as Bayesian, C4.5, k-mean, k- medoid and artificial neural network which based on special gloves. This method uses video and converts the video into frames to identify the pattern they know the hand gestures [5].

Recognizing of gesture language at present done by the token gesture to four languages (English, Arabic, Chinese and Persian languages) using video camera such as a professional camera, mobile camera, tablet camera, special camera, or laptop camera [6] then convert the video into images and also need to extracting the features of images at last classify each number to four languages (English, Arabic, Chinese and Persian languages) into corresponding voices, this paper focuses on how the gesture numbers of four languages (English, Arabic, Chinese and Persian languages) are translate into voices to make the dumb people communicate with other people through voices. The four languages (English, Arabic, Chinese and Persian languages) numbers gestures are shown in Figure 1.

2. Clustering algorithms and recognition

There are many algorithms used in this paper for recognition such as; Bayesian, C4.5, K-mean, K- Medoid algorithms and ANN. Which used for build database in training stage to compare the new images by the database stored from a training stage.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3. Proposed Algorithm

The proposed algorithm consists of two stages; training stage and testing stage, where the training stage contain five steps (images acquisition, pre-processing step, extraction of features, recognition and store the database) while the testing stage contains the same steps from 1-4 and then compare the features extracted from new image with features in database as shown in Figure2 we had published paper related to this work but in fewer details and one language.

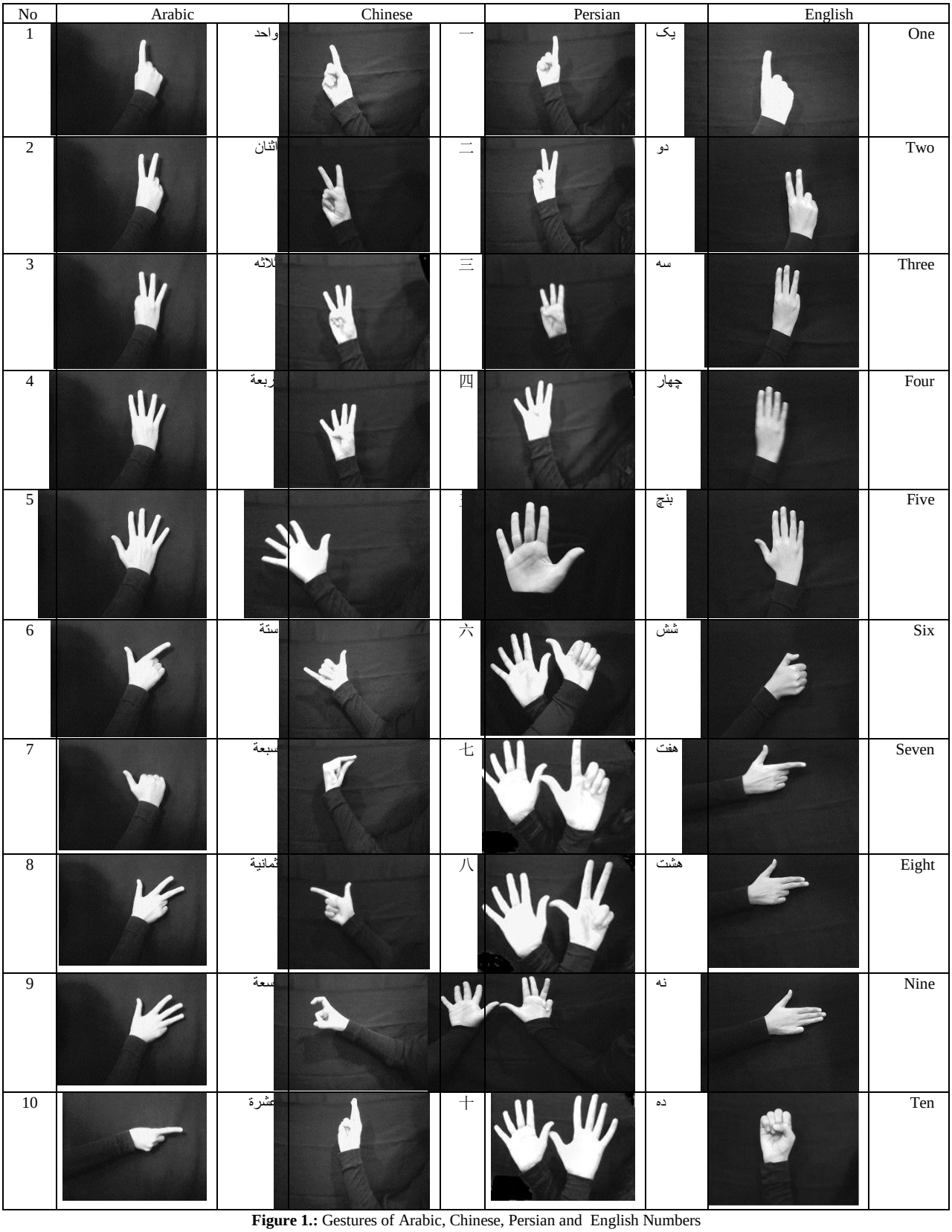


Figure 1.: Gestures of Arabic, Chinese, Persian and English Numbers

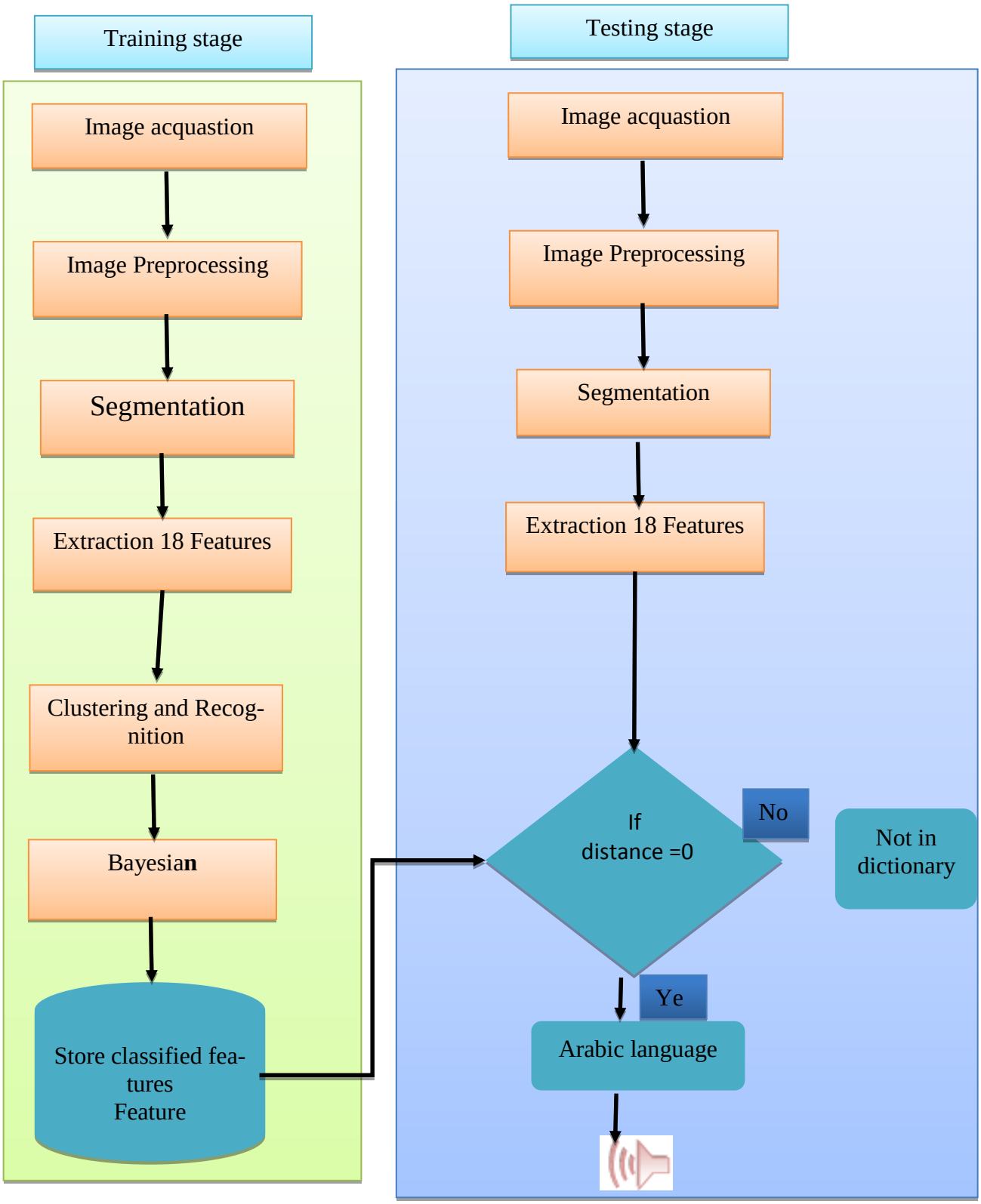


Figure 2: proposed algorithm

3.1 Dataset Acquisition

The database is created by using special cameras in the laboratory, where all images must have the same environments, the back-ground of images have black colour and fix the position of the camera to be easy to remove the background and keep the signal only In this paper there are eleven numbers of English, Arabic, Chinese and Persian (0, 1, 2, 3, 4, 5,6, 7, 8, 9, 10) from different persons, where these images will be templates for database. The images size (640 * 480 pixels) with JPEG formats.

3.2 Pre-Processing

The Pre-processing done by algrothim.1 the following steps as shown in Figure 3:

Algorithm 1

Input: image or frame from video.

Output: enhanced image.

1. Convert the video into images (frames).
2. Convert the image into gray scale format and resize image into 300*100 pixel.

3. Convert image into a binary image
4. Apply morphological operations to de-noise from binary image.
5. End.

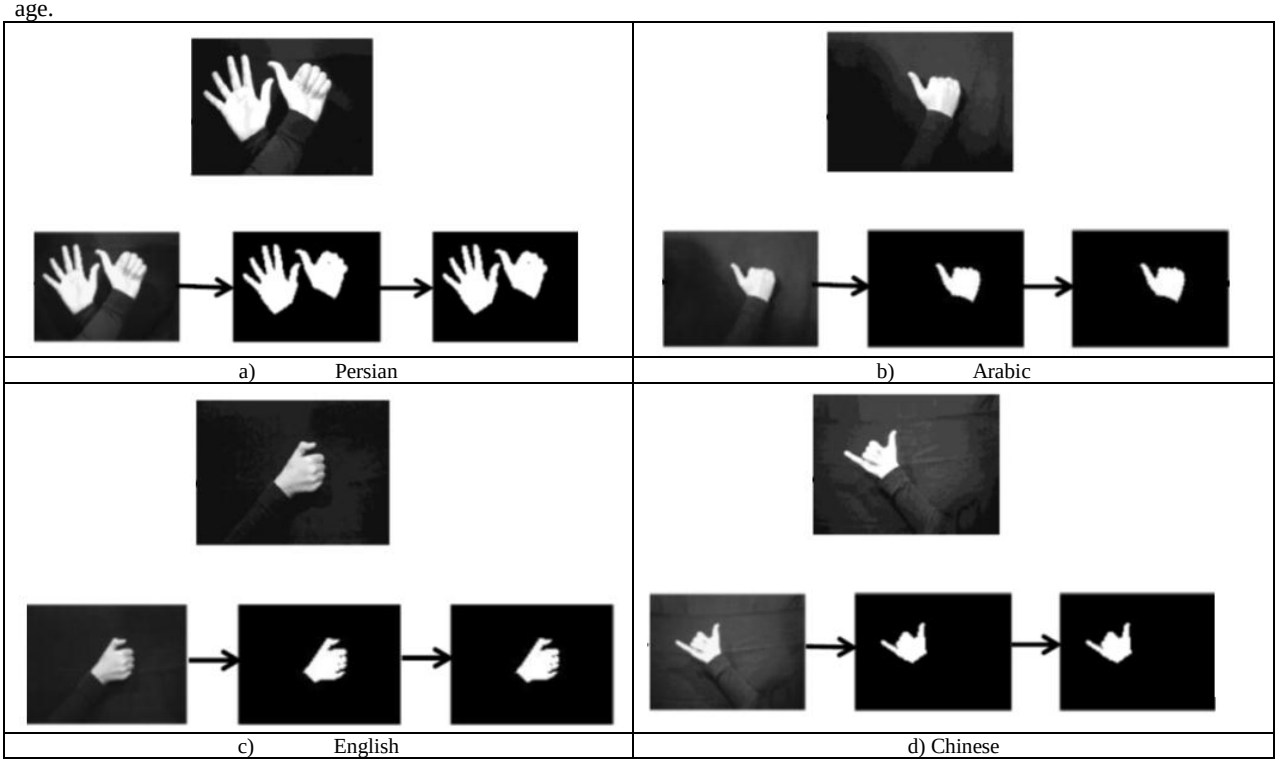


Figure 3: preprocessing steps for four languages

3.3 Segmentation

The Segmentation can be done by algorithm.2 as shown in Figure 4.

Algorithm 2

Input: enhanced image.

Output: binary image contains object (hand) only.

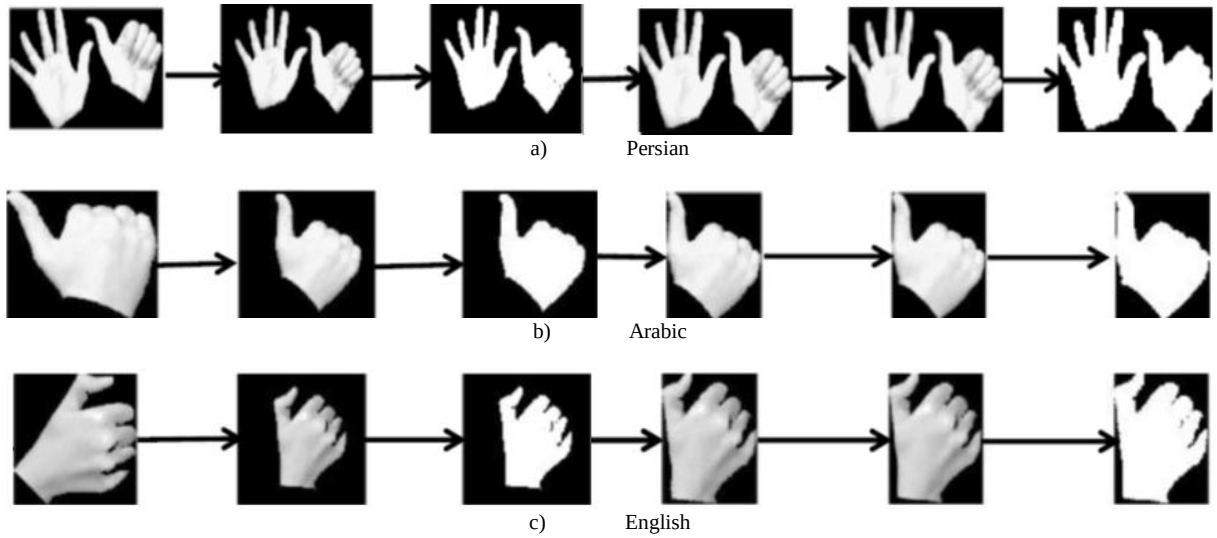
1. Create a mask from resulted image contains two values (1 and 0), 0's for the hand background and 1's for the hand sign part.
2. Multiply the resulted image from algrothm.1 by the mask.
3. Using equation (1) [8], to calculate the rotation angle of resulted image and rotate image by using equation (2) [9].

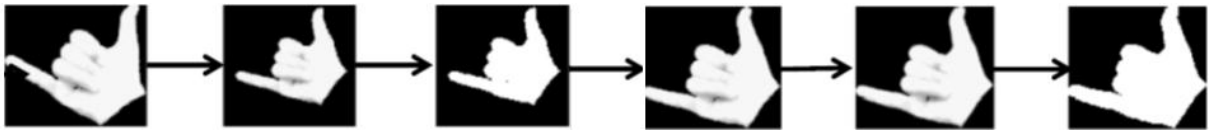
$$\theta = tan^{-1}\left(\frac{s_1 - s_2}{1 + s_1 * s_{2_1}}\right). \quad (1)$$

$where:\ s_1$ is the slope between A and B points is the s_2 slope between B and C points .

$$\left.\begin{aligned}\hat{z} &= z * cos(\theta) - w * sin(\theta)\\ \hat{w} &= z * sin(\theta) - w * cos(\theta)\end{aligned}\right\} \quad (2)$$

4. Create a new mask from rotated image contains two values (1 and 0) 0's for hand background and 1's for the hand sign part.
5. Multiply the resulted image from step 3 by a new mask and convert the image into a binary image.
6. End





d) Chinese

Figure 4: segment steps for four languages

3.4 Feature Extraction

In this paper, we used the geometric features where there are 18 geometrics features extracted from each input image, divided into (eight-features of distance between the eight points within the centre of hand by using equation (3)[7]) , (eight-features of slop between the eight points within the centre of hand by using equation (4)[8]) ,one- feature of area by using equation (6)[9] and last one-feature of perimeter (7)[10]. Also we need to find the centre of hand depending on equation (5)[11].

$$DE_{zo} = \sqrt{\sum_{i=1}^{n}(x_{zi} - x_{oi})^2} \qquad (3)$$

Where:

n: Number of properties, DE_{zo}: Distance between points and center of palm, x_{zi} : The coordinates of the i property for Z (points) , x_{oi}: The coordinates of the i property for o (where o: center point of palm)

$$Slop = \frac{yz - yo}{xz - xo} \qquad (4)$$

Where:

y_z: the y- axis points, x_z: the x- axis points, y_o: the y- axis of center value point of palm, x_o: the x- axis of center value point of palm.

$$x_o = \frac{\sum x_{oi} A_i}{\sum A_i}, \qquad y_o = \frac{\sum y_{oi} A_i}{\sum A_i} \qquad (5)$$

Where:

x_o: the x- axis of center value point of palm, y_o: the y- axis of center value point of palm, x_{oi}: The distance at which the center of the shape moves away from the junction point of the axes on the axis (x), y_{oi}: The distance at which the center of the shape moves away from the junction point of the axes on the axis (y), A_i: Area the shape.

$$A = \frac{1}{2}\sum_{i=0}^{n-1}(x_i \times y_{i+1}) - (x_{i+1} \times y_i) \qquad (6)$$

n: Number of points, x_i : x- axis coordinates points, y_i : y- axis coordinates points.

$$Per = \sum_{i=0}^{n-1} x_i \qquad (7)$$

n: Number of ribs, x: length of the rib.

This features can be extracted by using algorithm.3.

Algorithm 3

Input: binary image.

Output: features vector.

1. Calculate center point of palm as in equation (5).
2. Calculate the eight-features of distance by using equation (3) , then evaluate the length of the palm and divide the distance by the length.
3. Calculate the eight-features of slop from center of palm for eight-points as according to equation (4).
4. Calculate area of hand as according to equation (6).
5. Calculate Perimeter of hand as according to equation (7).
6. Calculate the feature vector for all numbers types (40 types) by find the average of features for the same type numbers from different images of the same number.
7. End

In testing stage the features will be compare with stored features where the input image will be classifying into 40 type of numbers and comparing with the feature vector with the 244 vectors stored in the training stage, calculated the distance in three methods; Standardized Euclidean distance by using equation (8), Euclidean distance and correlation by using equation (9).

$$`Dst = \sqrt{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{\sqrt{\frac{1}{(n-1)}\sum_{j=1}^{n}(x_j - \overline{x})^2 + (y_j - \overline{x})^2}}} \qquad (8)$$

Where:

x_i: is the ith valuse of first vector value, y_i: is the ith valuse of second vector value, n : is the number of elements in vector., $\overline{x}$: is the mean value of first and second vector.

$$R1 = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2}\sqrt{\sum_i (y_i - y_m)^2}} \qquad (9)$$

Where:

x_i is the intensity of the ith value in vector 1, y_i is the intensity of the ith value in vector 2, x_m is the mean intensity of vector 1, and my is the mean intensity of vector2.

4. Results

The proposed algorithm contains two stages; training stage using 244 images while the second stage (testing) there are 80 images used for testing represent four languages. There are 18 features for distance (D1, D2 …. D8), slop (S1, S2…. S8), area (A1) and perimeter (P1) respectively, then calculate the average for each type of number this will repeat for 40 type of numbers ten numbers for each language numbers as shown in Table.1. Table.2 shows the feature vector of the average image for all images in the database. The results of the five clustering algorithms; Bayesian, K-mean, K-mediod, C4.5, and ANN, for different numbers types (40 types) had high accuracy as shown in Table.3. and Figure. 5

In testing stage the last step is to convert the decision (a type of number in four languages) into corresponding voices of four lan-

guages such as five, بنج , 五 or خمسة , the first writing the same alphabet then play the sound that represents that number.

Table 1: Features vector of four languages numbers

Feature Geometry																		
Distance								Slope								A1	P1	Language
D1	D2	D3	D4	D5	D6	D7	D8	S1	S2	S3	S4	S5	S6	S7	S8			
0.1812	0.1877	0.1334	0.1341	0.1295	0.1255	0.2042	0.2013	-0.2210	-0.3426	1.4966	1.4470	0.4506	0.3787	-1.3742	-1.4718	0.3718	0.6753	chinses
0.2864	0.2735	0.1831	0.1814	0.2129	0.2123	0.1534	0.2825	0.5454	0.4619	-1.4347	-1.5625	0.1009	0.0696	1.4301	0.5677	0.7387	0.8324	Arabic
0.1207	0.1205	0.1019	0.1001	0.1038	0.1081	0.0986	0.1055	0.0776	-0.0607	-1.2688	-1.3324	-0.1225	-0.3072	1.2226	1.0734	0.1898	0.6542	Persian
0.2452	0.2407	0.1855	0.1824	0.2241	0.2205	0.1969	0.2330	0.2674	0.1872	-1.3729	1.4966	0.1880	0.0539	0.8512	0.6890	0.9008	1.1088	English

Table 2: Features vector for average of number six in four language

Feature Geometry																		
Distance								Slope								A1	P1	Language
D1	D2	D3	D4	D5	D6	D7	D8	S1	S2	S3	S4	S5	S6	S7	S8			
0.1812	0.1877	0.1334	0.1341	0.1295	0.1255	0.2042	0.2013	-0.2210	-0.3426	1.4966	1.4470	0.4506	0.3787	-1.3742	-1.4718	0.3718	0.6753	chinses
0.1207	0.1205	0.1019	0.1001	0.1038	0.1081	0.0986	0.1055	0.0776	-0.0607	-1.2688	-1.3324	-0.1225	-0.3072	1.2226	1.0734	0.1898	0.6542	Arabic
0.2864	0.2735	0.1831	0.1814	0.2129	0.2123	0.1534	0.2825	0.5454	0.4619	-1.4347	-1.5625	0.1009	0.0696	1.4301	0.5677	0.7387	0.8324	Persian
0.2452	0.2407	0.1855	0.1824	0.2241	0.2205	0.1969	0.2330	0.2674	0.1872	-1.3729	1.4966	0.1880	0.0539	0.8512	0.6890	0.9008	1.1088	English

Table 4: the accuracies of the four algorithms

Number	No image	Bayesian	C4.5	K-mean	K- mediod	ANN
	244	100%	100%	100%	100%	95%

As shown in Table .4, the Bayesian, C4.5, K-mean and K- mediod algorithm achieves the highest recognition accuracy100%, while ANN gave low than them 95%.

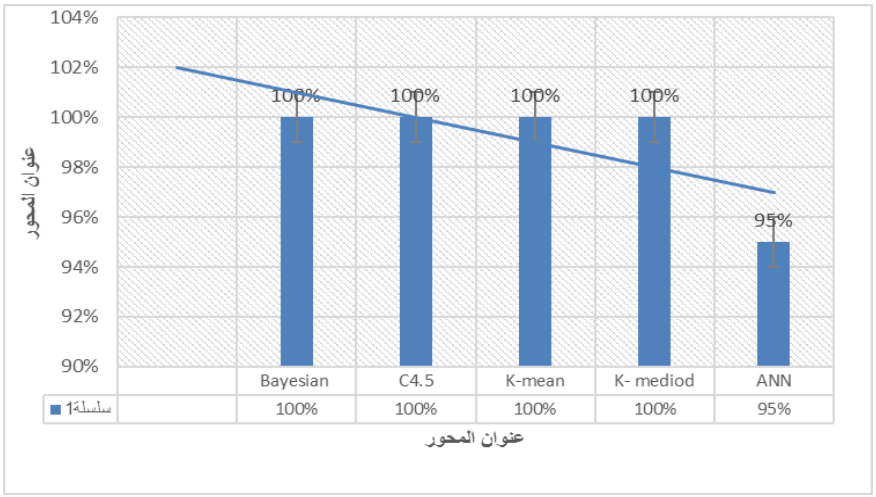


Figure 5: accuracy rate between five algorithms

5. Conclusion

In this paper, we have designed an algorithm for recognition four languages number based on clustering methods. The experiments show that the geometric features are best features than others features; shape, texture, colored…etc. These features (geometric features) achieved higher accuracy than the others. Classification algorithms used in our proposed algorithm such as Bayesian, K-

mean, K-medoid, C4.5, and ANN shows that Bayesian, K-mean, K-medoid, C4.5.5 algorithms are the best one in clustering or classifying with percentage of 100%, it is better than using other classification method such as GLCM which used in texture features and DWT which used with statistical features. In the testing stage, we found that the three methods; modified Standardized Euclidean distance, Euclidean distance and Correlation are the best metrics for recognition the 40 numbers.

References

[1] Ch. V.N. Syam Babu, V.J.K. Kishor Sonti and Y. Varthamanan ," Design and Simulation of Communication Aid for Disabled Using Threshold Based Segmentation" , International Journal of Control Theory and Applications 9(7), pp. 3275-3281, 2016.

[2] Mohamed S. Abdalla, Elsayed E. Hemayed," Dynamic Hand Gesture Recognition of Arabic Sign Language using Hand Motion Trajectory Features", Global Journal of Computer Science and Technology Graphics & Vision, Volume 13, Issue 5,p. 26-33,2013.

[3] Hemina Bhavsar, Dr. Jeegar Trivedi," Review on Classification Methods used in Image based Sign Language Recognition System", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 5, pp. 949 – 959,2017.

[4] Kumud Tripathi, Neha Baranwal and G. C. Nandi,"Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Procedia Computer Science ,54, pp. 523 – 531,2015.

[5] Liu Yun, Zhang Lifeng and Zhang Shujun, "A Hand Gesture Recognition Method Based on Multi-Feature Fusion and Template Matching", Procedia Engineering, 29, pp. 1678 – 1684, 2012.

[6] Pablo Barros, Nestor T. Maciel-Junior , BrunoJ.T. Fernandes , ByronL.D. Bezerra and Sergio M.M. Fernandes ," A dynamic gesture recognition and prediction system using the convexity approach",Computer Vision and Image Understanding,155,pp.139-149,2017.

[7] Michel Marie Deza , Elena Deza," Encyclopedia of Distances", Springer, p. 94,2009.

[8] Christopher Clapham, James Nicholson, " Oxford Concise Dictionary of Mathematics", OUP oxford,2009.

[9] Michel Marie Deza , Elena Deza," Encyclopedia of Distances", Springer, p. 94,2009.

[10] RAMKUMAR, C., KUMAR, H. P., & KRISHNAPRASANTH, S. EAACM: Enhanced ACK Aware Clustering Mechanism for Energy Efficient and Secure Routing in Wireless Sensor Networks.

[11] Dr Yeap Ban Har,Dr Joseph Yeo,Teh Keng Seng,Loh Cheng Yee,Ivy Chow,Neo Chai Meng and Jacinth Liew , "New Syllabus Mathematics Teacher's Resource Book1",7th edition,Oxford University Press,2018.

[12] Dan B. Marghitu , Mihai Dupac, " Advanced Dynamics",ch2, Springer, New York, NY,p.73-141,2012.