AUTOMATIC ISOLATED-WORD ARABIC SIGN LANGUAGE RECOGNITION SYSTEM BASED ON TIME DELAY NEURAL NETWORKS: NEW IMPROVEMENTS

[1] FERAS FARES AL MASHAGBA, [2] EMAN FARES AL MASHAGBA , [3] MOHAMMAD OTHMAN NASSAR

[2]*Asstt Prof., Department of computer information systems, Zarqa University, Zarqa, Jordan*

[1,3] Asstt. Prof., Department of computer information systems, Amman Arab University, Amman, Jordan

E-mail: [1] ferasfm79@yahoo.com, [2] Emanfa71@yahoo.com , [3] moanassar@yahoo.com ,

ABSTRACT

This research presents an improved automatic isolated-word recognition system for the Arabic sign language (ArSL) for the Jordanian accent. Our proposed system requires that the signer wears two gloves with different colors, he also should wear another colored mark on his head, this mark should have different color than the colors used in the gloves. In this paper the video for each sign is converted to a sequence of static images; each image is segmented for three colored regions and outlier according to the mean and covariance of each color region using the multivariate Gaussian Mixture Model (GMM) on the characteristics of the the Hue Saturation Lightness Model (HSL) color space. Tracking the hand motion trajectories of the right and left hand over time is conducted . Finally we identify a list of features to be used as an input to the time delay neural networks (TDNN) for the recognition step. Two different test collections were used in this research, the first data collection is used to prove that when using the head of the signer to determine the position of the centroid for the right and left hand instead of using the center of the upper area for each frame as reference can improve the recognition rate. The experimental results shows an improvement in the recognition rate from 70% to 71.66%. finally we used the second data collection to prove that our proposed categorization for the Arabic signs into four categories according to the overlap existence between head and hands can improve the recognition rate. The experimental results based on data collection number two shows an improvement in the recognition rate for the testing set, where the recognition rate increased to reach 77.43%.

Keywords: *Sign Language Recognition Systems; Isolated-Word based system; video based system; TDNN; Arabic signs; Jordan accent.*

1.   INTRODUCTION

Sign language is used for communication between hearing–impaired people and hearing people. Creating an Automated sign language recognition system can be considered as a translator between the hearing-impaired individuals and the ordinary people. Usually the isolated sign language recognition systems takes one word as input, isolated sign language recognition systems are considered one of the most important topics in sign language recognition since they are designed to process the main basic unit in the sign language [4, 6]. Sign language recognition can be classified into two categories; vision-based and device-based. vision-based systems depends on using video camera to capture the signs. This approach have lower accuracy and needs high computing power

consumption [7]. In the device-based systems, the user wear some device with sensors that help in capturing the physical features of the signer gestures. This approach needs complex devices and sensors that may cost too much.

differences in sign language might be found because signs are mostly created by hearing-impaired individuals themselves [5]. Limited research has been conducted based on ArSL [2,6], so the Arabic Sign Language Jordanian accent is used in this research.

Designing an automated Arabic Sign Language recognition systems attempts can be summarized into two categories; Static image based Arabic Sign Language recognition systems such as [2,9,12,13] and Video based Arabic Sign Language recognition systems such as[10,11,6]. In [2] the authors creates a vision based system to

recognize the static gestures of alphabets with recognition accuracy of 90.55 %. In [9] and based on Hidden Markov Model (HMM) the authors created an image based system for Arabic sign language with recognition accuracy of 98%. In [13] different types of neural networks were used to recognize static images for human hand gestures as well as for dynamic gestures, they have used different architectures for feed forward neural networks and recurrent neural networks. Our proposed system will use dynamic gestures (Video) that represent Arabic words since most of the previous work dialed with static images for the Arabic alphabets [2,12].

In continuous sentence recognition systems the system can recognize more than one word, While the isolated-word Arabic Sign Language recognition system can deal with one word at a time. In [1] the authors introduced the first automatic isolated-word ArSL recognition system based on HMMs, unlike the proposed system in this research the proposed system in [1] did not rely on input devices such as gloves . sensor-based gloves system for Arabic sign language recognition is introduced by the authors in [10], the system used an extraction method based on accumulated differences (ADs). Their proposed system yielded recognition rate of 92.5% for the user dependent model, and 95.1% for the user independent model. In [8] 3D-views model for the Arabic hand postures recognition is introduced, the researchers used Pulse Coupled Neural Network. The authors in [11] implement a real-time Arabic Sign Language Recognition System to recognize real-time connected sequence of gestures, the system is based on pulse-coupled neural network.

Time Delay Neural Network Algorithm (TDNN) have been used for the first time by [6] to recognize the isolated-word Arabic Signs for the Jordanian sign language with recognition rate of 70%, the authors used the TDNN because of its high ability in learning spatial-temporal patterns. The authors in [6] creates their own data set that is composed from 40 isolated words from 6 different domains coded by 10 different signers. The researchers in [6] used the center of the upper area for each frame as a reference to determine the position of the centroid for the right and left hand. Unlike the system proposed by [6] and in order to improve the recognition rate; we are going to use the head of the signer to determine the position of the centroid for the right and left hand.

This research is designed to improve the recognition rate for the proposed system by [6]. The improvement will be achieved using three important modifications. The first modification is to use the head of the signer to determine the position of the centroid for the right and left hand instead of using the center of the upper area for each frame as reference, this modification is proposed due the fact that we notice that the body of the signer including his both hands usually are moving together in defined patterns with the signer head, we notice that these patterns are different form signer to signer, so by using the center of the upper area for each frame as references we lose those patterns leading to low recognition rate.

The second modification is the addition of additional color to the signer head , this will help in two directions; the first benefit is to easily determine the head using the color. The second benefit is to give us the ability to recognize the words that the head is usually overlap with the hands in their gesture.

The third modification is the categorization of the sign language words into four new categories. the authors in [6] used sign language words from six domains, the domains was created according to the domain area, where each area contains words that are related to that area. in order to improve the recognition rate we created four new categories based on our analysis for the gestures, we notice that the signer represents some words gestures using his head and his two hands, sometimes he represent words using only one hand, other times he use one hand and his head, so we decide to categorize the sign language words into four categories ; Hand with head, Hand without head, Two Hand with head, Two Hand without head.

2. MATERIALS AND METHODS

We will use the same system parameters as mentioned by [6] to allow us to compare our results to their results. We used video camera to capture the signer upper body. Each signer wear colored gloves with yellow for the Right hand, and blue for the left hand. The face of the signer is tied up by red color. Each captured video has 80 to 220 frames with image size of 243×360 pixels, 24bit color in TIFF format. The lighting and background in each sequence is held constant.

We choose two data collections to test the proposed system. The first data collection is the

same data collection used by [6], in that data collection 40 Arabic Signs shown in table 1 are chosen then categorized according to the subject into six domains, table 1 can be found at the end of the paper, each word in this collection is signed by 10 different signers. The second data collection is created by us, this

collection is composed from 12 words shown in table 2 divided into four categories, the categorization is based on the overlap existence between the hands and the head, each word in the second data collection is signed by 10 different signers.

Table 2: Arabic Sign Language Signs Categorized

According To The Overlap Existence.

Category Name	ArSL Word	English Gloss
Hand with head	أب	Father
	السلام عليكم	Peace be upon you
	أم	Mother
Hand without head	دكتور	Doctor
	شارع	Street
	طائره	Aeroplane
Two Hand with head	أخت	Sister
	تلفاز	Television
	شكرا	Thanks
Two Hand without head	أهلا وسهلا	Welcome
	حليب	Milk
	هديه	Gift

We Convert each sign video to sequence of images (frames), for each frame we used the Gesture Mixture Model (GMM) to separate and classify all color pixels. Then the three colors (yellow for the right hand region , blue for the left hand, and red for the head) were extracted using the MATLAB function roipoly. at this stage the image is identified in three colored regions and outlier according the the mean and covariance for each color.

For a given video sequence and to track the hand and head motion trajectories over time, we made a list for the positions of the centroid for each of the right, and left hand , and face in each frame. We also made a list of features, such as the angle of velocity for each hand, the centroid position for each hand using the head as reference, the horizontal and vertical velocity of both hands across two consequence frames using change in position over time, and the area change for each hand across the two frames. a vector $yi= [xi, yi, vi, vj, \theta i]$ where created for each frame, where xi , yi are position of both hands with respect to center of face , vi, vj the horizontal and vertical velocity of both hands across two frames using change in position over time, and θi angle of velocity of both hands . finally all frame vectors are stacked to each other to form a feature vector for motion trajectory which used as input to Time Delay Neural Network (TDNN) to recognize the gesture. We decide to choose the Hyperbolic tangent sigmoid transfer function 'tansig' $(a = tansig(n) = 2/(1+exp(-2*n))-1)$ as our activation functions in the hidden layer for the TDNN. In the output layer we used the linear transfer function 'purelin' $(a=purelin(n)=n)$.

We used the Levenberg-Marquardt (LM) back propagation training function to updates weight and bias values according to Levenberg-Marquardt (LM) optimization. To terminate the learning process we used the same conditions as stated by [6], in [6] the authors propose four conditions to terminate the learning process; first: The maximum number of repetitions is reached. Second: The maximum amount of time has been exceeded. Third: the Performance has been minimized to the goal. Forth: the performance gradient falls below minimum. if anyone from these four conditions is met the learning process is immediately terminated. In this research the sum square error is calculated according to the difference between the target and the output of the neural network and used as the performance

criteria for our improved ArSL recognition system.

3. RESULTS AND DISCUSSION

We used the MATLAB Toolbox to implement the image processing and TDNN code. In this experiment; five features (called motion trajectories) have been chosen to represent each hand features, they are $(x_i, y_i, v_i, v_j, \theta_i)$. Where θ_i is the Angle of velocity for the hand, v_i, v_j are the Magnitudes of horizontal and vertical velocity for the hand across two consequence frames using change in position over time, x_i, y_i are the position of the hand with respect to the center of head.

The hand motion segmentation method used here is identical to the method presented in [6], it consists of two major steps; First: each image (frame) is partitioned into three colored regions (yellow, blue, and red) and one outlier using a color cue, then the hand locations are specified with reference to a head region. Second: the motion trajectories $(x_i, y_i, v_i, v_j, \theta_i)$ of each hand over time (from frame to frame) is calculated.

The total number of features for both hands will be ten features, those features will be used as an input to the TDNN to classify gestural motion patterns. The authors in [6] conducted Many experiments to choose the appropriate network topology, they did choose the topology presented in Table 3. For comparison resons we are going to choose the same topology as stated by [6].

Table: 3 Neural Network Architecture

Layers	Nodes	Activation Function	Number of Delays
Input layer	10 Nodes	Tansig	4
Hidden layer one	22 Nodes	Tansig	4
Hidden layer two	22 Nodes	Tansig	6
Output layer	1 Nodes	Purlin	6

To test our improved system compared to the system proposed by [6], we should use the same data collection that have been used by [6], this collection is described earlier in this paper, for each domain in table 1 we take 50% for training

the TDNN and the other 50% are kept for testing. Table 4 shows the results for both systems. the results clearly shows that using the head of the signer as a reference instead of using the upper area as a reference can improve the recognition rate. This can be explained by our observation which argues that the signer hands usually are moving together in defined patterns with the signer head, we notice that these patterns are different form signer to signer, so by using the center of the upper area for each frame as references we lose those patterns leading to low recognition rate.

Table 4: Recognition Rate For The Old And The Improved Systems.

Category Name	Recognition Rate in Testing Set for the improved system.	Recognition Rate in Testing Set for the old system.
Domain number 1	74.1%	72.2%
Domain number 2	69.2%	67.7%
Domain number 3	73%	71.1%
Domain number 4	72.3%	69.9%
Domain number 5	66%	65.5%
Domain number 6	75.4%	73.3%
Total recognition rate	71.66%	70.0%

The second data set which is created and described by us earlier in this paper will be used

to prove our second argument, the argument states that categorizing the Arabic signs according to the existence of overlap between the head and the hands will increase the system ability to recognize the gestures. Table 5 shows that the total recognition rate for our improved system exceeded the previous recognition rates in table 4.

Table 5: Recognition Rate In Training And Testing Set For The Data Collection Number Two.

Category Name	Recognition Rate in Training Set	Recognition Rate in Testing Set
Hand with head	100%	75.5
Hand without head	100%	86.6%
Two Hand with head	100%	78.8%
Two Hand without head	100%	68.8%
Total recognition rate	100%	77.43

4. CONCLUSIONS

Our research proposed new ideas to improve an existing automatic isolated-word Arabic sign language recognition system based on time delay neural networks. Our proposed ideas were tested against two different test collections. The experimental results shows that our proposed ideas were able to achieve an improvement in the recognition rate for the testing sets. For future work we recommend creating and using larger data collections for the Arabic sign language, this will help in validating the results of this research and other researches in the literature regarding the Arabic signs

REFRENCES:

[1] AL-Rousan M., Assaleh K. and Tala'a A., "Videobased signer-independent Arabic sign language recognition using hidden Markov models", journal of Applied Soft Computing, Vol. 9 Issue 3, June 2009, pp. 990-999.

[2] Nadia R. Albelwi , Yasser M. Alginahi , "Real-Time Arabic Sign Language (ArSL) Recognition", International Conference on Communications and Information Technology (ICCIT-2012), Tunisia, 2012, pp.497-501.

[3] Y. Cui, J. Weng , "Hand Segmentation Using Learning-Based Prediction and Verification for Hand Sign Recognition ", (CVPR'96), IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 88-95, 1996.

[4]] Sandjaja I., Marcos N., "Sign Language Number Recognition," Fifth International Joint Conference on INC, IMS and IDC, pp.1503-1508, 2009.

[5] T. Shanableh1, K. Assaleh2 and M. Al-rousan, "Spatio-Temporal feature extraction techniques for isolated gesture recognition in Arabic sign language", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol 37(3), 2007, pp 641 - 650.

[6] Feras Fares Al Mashagba, Eman Fares Al Mashagba, mohammad Othman Nassar," Automatic Isolated-Word Arabic Sign Language Recognition System Based On Time Delay Neural Networks", accepted for publication in Research Journal of Applied Sciences, Engineering and Technology. 2013.

[7] Charayaphan C. and Marble A., "Image processing system for interpreting motion in American SignLanguage," Journal of Biomedical Engineering, Vol. 14, 1992, pp.419-425.

[8] Samir, Ahmed; Abul-Ela, Magdy,"3D Arabic sign language recognition using linear combination of multiple 2D views.", 8th International Conference on Informatics and Systems(INFOS, 2012), pp. 6-13.

[9] Mohandes, M. and Deriche, M., "Image based arabic sign language recognition.", Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 2005, pp. 86- 89.

[10] Khaled Assaleh, Tamer Shanableh, Mohammed Zourob, "Low Complexity Classification System for Glove-Based Arabic Sign Language Recognition.", Neural Information Processing: Lecture Notes in Computer Science, Volume 7665, 2012, pp 262-268.

[11] M. F. Tolba, Ahmed Samir, Magdy Aboul-Ela, "Arabic sign language continuous sentences recognition using PCNN and graph matching", Neural Computing and Applications journal. August 2012, pp.1-12.

[12] Reyadh Naoum, Hussein H. Owaied, Shaimaa Joudeh, "Development of a New Arabic Sign Language Recognition Using K-Nearest Neighbor Algorithm.", Journal of Emerging Trends in Computing and Information Sciences. VOL. 3, NO. 8 August, 2012.

[13] Manar Maraqa, Farid Al-Zboun, Mufleh Dhyabat and Raed Abu Zitar, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks.", Journal of Intelligent Learning Systems and Applications, Vol:4, pp.41-52, 2012.

Table 1 Standard Arabic Sign Language Signs Categorized According To The Subject [6].

Sign No.	Arabic Sign (word/phrase)	English Meaning	Sign No.	Arabic Sign (word/phrase)	English Meaning
Domain 1: Adjectives and Feelings			Domain 4: Money and Commerce		
1	سعيد	Happy	22	نقود	Money
2	مهم	Important	23	ربح	Profit
3	عصبي	Nervous	24	مصرف	Bank
4	جميل	Beautiful	25	دينار	1 JD
5	حب	Love	26	مجانا	Free of Charge
6	قلق	Worry	Domain 5: Hospital		
7	جاد	Serious	27	مستشفى	Hospital
8	جديد	New	28	طبيب	Doctor
Domain 2: Home and Visits			29	مريض	Patient
9	جار	Neighbor	30	صداع	Headache
10	صديق	Friend	31	دواء	Medication
11	هدية	Gift	32	حقنة	Injection
12	اهلا وسهلا	Welcome	33	فحص	Laboratory Examination
13	بيت	Home	34	دم	Blood
14	كيف حالك	How are you	35	عملية جراحية	Surgical Operation
15	السلام عليكم	Hello	Domain 6: Miscellaneous		
16	مشاكل	Problem(s)	36	طعام-يأكل	Food / eat
Domain 3: Roads and Transportation			37	ماء-يشرب	Water/ drink
17	حافلة	Bus	38	كم	How many/much
18	تكسي	Taxi	39	اين	Where
19	محطة	Station	40	لماذا-السبب	Why /because
20	أجرة	Fare			
21	شارع	Street			