Research Journal of Applied Sciences, Engineering and Technology 7(11): 2261-2265, 2014
DOI:10.19026/rjaset.7.524
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: June 17, 2013 Accepted: July 12, 2013 Published: March 20, 2014

Research Article

Automatic Isolated-Word Arabic Sign Language Recognition System Based on Time Delay Neural Networks

[1]Feras Fares Al Mashagba, [2]Eman Fares Al Mashagba and [1]Mohammad Othman Nassar
[1]Amman Arab University, Amman, Jordan
[2]Zarqa Private University, Zarqa, Jordan

Abstract: There have been a little number of attempts to develop an Arabic sign recognition system that can be used as a communication means between hearing-impaired and other people. This study introduces the first automatic isolated-word Arabic Sign Language (ArSL) recognition system based on Time Delay Neural Networks (TDNN). The proposed vision-based recognition system that the user wears two simple but different colors gloves when performing the signs in the data sets within this study. The two colored regions are recognized and highlighted within each frame in the video to help in recognizing the signs. This research uses the multivariate Gaussian Mixture Model (GMM) based on the characteristics of the well known Hue Saturation Lightness Model (HIS) in determining the colors within the video frames. In this research the mean and covariance of the three colored region within the frames are determined and used to help us in segmenting each frame (picture) into two colored regions and outlier region. Finally we propose, create and use the following four features as an input to the TDNN; the centroid position for each hand using the center of the upper area for each frame as references, the change in horizontal velocity of both hands across the frames, the change in vertical velocity of both hands across the frames and the area change for each hand across the frames. A large set of samples has been used to recognize 40 isolated words coded by 10 different signers from the Standard Arabic sign language signs. Our proposed system obtains a word recognition rate of 70.0% in testing set.

Keywords: Arabic signs, isolated-word based system, Jordan accent, sign language recognition systems, TDNN, video based system

INTRODUCTION

Sign language is considered as the only way used for communication between hearing-impaired people and hearing people. In fact differences in sign language might be found among speakers in the same country and sometimes even among speakers in neighboring cities. This is because signs are mostly created by hearing-impaired individuals themselves and are highly influenced by the local environment (Shanableh *et al*., 2007). Automated sign language recognition system would make deaf-hearing interaction easier, particularly public functions such as the courtroom, conversions and meetings.

According to (Sandjaja and Marcos, 2009); the three key types for sign language recognition are: hand shape, isolated sign and continuous sign. The major basic units in the sign language are the isolated signs; and that's why they are considered as a key and important subject in the sign language recognition (Sandjaja and Marcos, 2009).

Sign language recognition approaches can be classified into two major parts: first; the vision-based. In this approach we usually needs a complex image processing system that takes the inputs from a camera to distinguish and categorize the signs with lower accuracy if compared to the device-based approach (Charayaphan and Marble, 1992). Second; the device-based approach. In this approach the signer have to wear some kind of devices with wide range of sensors types to help in the recognition process for the signs by extracting the physical features of those signs such as angles and the dimensions for the sign motions using the sensors (Kramer and Leifer, 1988).

The Arabic Sign Language accent used in this research is basically the Jordanian accent. Compared to other sign languages, not much has been done in the automation of Arabic Sign Language (ArSL). Arabic Sign Language is the language of choice for most deaf people in more than 20 Arabic countries covering a large geographical and demographical portion of the world. ArSL has received little attention in Sign Language research (Nadia and Yasser, 2012) except for few attempts which will be discussed later in the Previous Studies section.

In this study a new approach is used to develop an automatic isolated-word Arabic Sign Language recognition system for the Jordanian sign language. The

Corresponding Author: Feras Fares Al Mashagba, Amman Arab University, Amman, Jordan

proposed system is the first Arabic recognition system which based on Time Delay Neural Network algorithm (TDNN). The proposed system is the first system for the ArSL isolated-word that uses the idea of colored gloves to aid the recognition process. In this research we propose, create and use our own set of features to use them as an input to the TDNN. This research uses vision-based recognition principles to develop an automatic isolated-word Arabic Sign Language recognition system with the help of an image processing approaches and ideas.

There was a little number of attempts to design an automated Arabic Sign Language recognition system. These attempts can be summarized according to the type of the input data to the recognition system into two categories: first; Static image based Arabic Sign Language recognition systems such as Nadia and Yasser (2012), Mohandes and Deriche (2005), Reyadh *et al*. (2012) and Manar *et al*. (2012). Second; Video based Arabic Sign Language recognition systems that can deal with dynamic gestures (Video) for the Arabic signs such as Khaled *et al*. (2012) and Tolba *et al*. (2012). In Nadia and Yasser (2012) the authors presents a vision based system that provides a solution to Arabic Sign Language (ArSL) recognition of static gestures of alphabets. The proposed method doesn't require that signers wear gloves or any other marker. Their system achieved a recognition accuracy of 90.55%. In Mohandes and Deriche (2005) the authors propose an image based system for Arabic sign language recognition using Hidden Markov Model. Their proposed system achieves a recognition accuracy 98% for a data set of 50 signs. In Manar *et al*. (2012) the authors introduced the use of different types of neural networks in human hand gesture recognition for static images as well as for dynamic gestures. This study focuses on the ability of neural networks to assist in Arabic Sign Language (ArSL) hand gesture recognition. They have presented the use of feed forward neural networks and recurrent neural networks along with its different architectures. Our proposed system will deal with dynamic gestures (Video) since most of the previous work dialed with static images for the Arabic alphabets (Nadia and Yasser, 2012; Manar *et al*., 2012).

Video based Arabic Sign Language recognition systems can be further categorized into two main types: first; continuous sentences recognition systems that can recognize more than one sign such as Tolba *et al*. (2012). Second; isolated-word Arabic Sign Language recognition system such as AL-Rousan *et al*. (2009) and Khaled *et al*. (2012). AL-Rousan *et al*. (2009) introduced the first automatic ArSL recognition system based on HMMs in 2009. They used large set of samples to recognize 30 isolated words from the standard ArSL. Their system did not rely on input devices such as gloves. In Khaled *et al*. (2012) the authors presented a system for Arabic sign language recognition using sensor-based gloves. Their system used a classification approach based on a novel feature extraction method based on Accumulated Differences (ADs). Their proposed system is applied to a dataset of Arabic sign language gestures and it yielded a recognition rates 92.5 and 95.1% for user dependent and user independent models, respectively.

In Samir and Abul-Ela (2012) a new model is proposed and developed for three dimensional views, the model was used in Arabic hand postures recognition. The researchers used Pulse Coupled Neural Network to create the features vector for single view. Using other angels the researchers created another 2 views; for each view they create features vector. Then the authors linearly combined the two dimensional vectors to construct there dimensional features which finally have been used for the recognition process. The problem of Arabic sign language recognition for continuous sentences has been studied by Tolba *et al*. (2012). They implement a real-time Arabic Sign Language Recognition System which applied pulse-coupled neural network to recognize the real-time connected sequence of gestures. Our proposed system will deal with isolated-word since the isolated-words are the main basic unit in the sign language (Sandjaja and Marcos, 2009). Our proposed system will be the first system for the ArSL isolated-word that uses the idea of colored gloves to aid the recognition process.

Artificial neural networks have been used in Arabic sign language recognition research such as Samir and Abul-Ela (2012), Tolba *et al*. (2012) and Manar *et al*. (2012). None of these previous studies used the Time Delay Neural Network algorithm (TDNN). We are going to use the TDNN in this research to solve the problem of isolated-word Arabic Sign Language recognition for the Jordanian Arabic sign language. TDNN have been demonstrated to be very successful in learning spatial-temporal patterns.

MATERIALS AND METHODS

The proposed ArSL recognition system is built as following:

Capturing the video: The video camera captures the sign word conducted by the signers with the following Assumptions:

- The view is chosen so that the signer upper body is captured
- The subject wear colored gloves, where yellow color for Right hand and blue color for left hand
- The subject wear colored gloves, unlike background and clothing

Word selection: Forty Arabic Signs are chosen then categorized according to the subject into six domains.

Convert video to sequence of images: The captured video segments are converted into image frames. Now each sign is represented by a sequence of frames.

Segmentation: The segmentation step aims to separate and classify all color pixels in connected components. In this research Gesture Mixture Model (GMM) is used.

Region of Interest (ROI): The mask of the yellow right hand region and blue left hand was taken by using MATLAB function roipoly. This method was chosen so that user can have freedom to select any desirable part of an image. Since mask consists of one's within enclosed area selected by the user and zeros elsewhere. Accumulate the data for a set of masks from their corresponding images and return the right hand and the lift hand masks, then calculated the mean and covariance for each colored region. Segmentation image to two colored regions and outlier according the mean and covariance.

Geometric analysis: Color segmentation partitioned each frame to right hand with yellow color, left hand with blue color and outlier region with another different color. For a given video sequence, we made a list of the positions of the centroid for each of the right hand and left hand in each frame. Tracking the hand motion trajectories of the right hand and left hand over time was fairly simple. We also made a list of features, the position of the centroid for each of the right hand and left hand using the center of the upper area for each frame as references, the horizontal and vertical velocity of both hands across the two frames using change in position over time. And the area change for each hand across the two frames. For each frame we form a vector yi = [xi, yi, vi, vj, Ai], where xi, yi are position of both hands with respect to the center of the upper area for each frame, vi, vj the horizontal and vertical velocity of both hands across the two frames using change in position over time and Ai area change for each hand across the two frames. All frame vectors are stacked to each other to form a feature vector for motion trajectory which is used as input to classification stage to recognize the gesture.

Time Delay Neural Network (TDNN): We will use the TDNN in the process of classifying the gesture motion patterns. The previous researches showed that the TDNNs are very successful in learning spatio-temporal patterns and thus we are going to use them in this research. The nature of the networks used in developing the ArSLR system is as following:

- **Activation functions:** During this research a Hyperbolic tangent sigmoid transfer function 'tansig' function was used in hidden layer. Hyperbolic tangent sigmoid transfer function:

  a = tansig (n) = 2/ (1 + exp (-2*n)) - 1

  Linear transfer function purelin is used in output layer. To calculate a layer's output from its net input:

  a = purelin (n) = n

- **Networks architecture:** The determination of a suitable topology depends greatly on the experience of the person controlling the network's training. TDNN architecture is used in this research.
- **Input and output encodings:** The raw data to be processed by the network has to be encoded and placed into the input nodes of the network. If the input values are numeric then the encoding is straightforward, as the data values can be placed directly into the input nodes. The usual practice (also adopted during this research) is to scale these input values so that they all lie in the same range, as this can aid the speed and success of training methods.
- **Learning algorithms:** This research proposes to use Levenberg-Marquardt (LM) backpropagation training function that updates weight and bias values according to Levenberg-Marquardt (LM) optimization.
- **Length of training:** Algorithm stops when any of these conditions reached:

  - The maximum number of repetitions
  - The maximum amount of time
  - The Performance minimized to the goal

Evaluation: Training the TDNN and then perform the test to find the recognition rate.

RESULTS AND DISCUSSION

Unlike other sign languages, Arabic dose not yet have a standard database that can be purchased or publicly accessed. Therefore, we collect our own ArSL video database of 40 signs categorized into 6 domains as shown in Table 1, this table is taken from (AL-Rousan *et al*., 2007). The 40 signs were filmed using Panasonic camera by 10 signers. Each video consists of an ArSL sign with image size of 243×360 pixels, 24 bit color in TIFF format. We keep the lighting and background conditions constant during the experiment to remove their effect on the results since they are going to be constant amongst all the captured videos. An image sequence of each of the 40 gestures in the experiments has 80 to 220 frames. The signs vocabulary used in our system is shown in Table 1. The MATLAB Toolbox was used to implement the image processing and TDNN code.

The chosen features describe the two-dimensional projection at each hand in the image plane are:

- Positions with respect to the center of the upper area for each frame (x, y)
- Magnitudes of velocity (vi, vj)
- Area change for each hand across the two frames (Ai)

Table 1: Standard Arabic sign language signs

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	1JD
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			

AL-Rousan *et al*. (2007)

Table 2: 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

Table 3: Recognition rate in training and testing set

Category name	Recognition rate in training set (%)	Recognition rate in testing set (%)
Domain number 1	100	72.2
Domain number 2	100	67.7
Domain number 3	100	71.1
Domain number 4	100	69.9
Domain number 5	100	65.5
Domain number 6	100	73.3
Total recognition rate	100	70.0

The hand motion segmentation method used here consists of two major steps First, each image is partitioned into regions using a color cue, where right hand has a yellow color and left hand has a blue color are extracted from each frame and hand locations are specified with reference to the center of the upper area for each frame. It should be emphasized that this motion segmentation algorithm is depending on color and geometric information; second, the motion trajectories (xi, yi, vi, vj, Ai) of each hand over time is calculated. In classification stage we employ TDNN to classify gestural motion patterns of hand regions since TDNNs have been demonstrated to be very successful in learning spatial-temporal patterns. TDNN can be used effectively to develop general feature detectors by manipulating the weights over the nodes. Many of experiments were conducted to choose the appropriate network topology as shown in Table 2. The inputs to our TDNN are vectors of (xi, yi, vi, vj, Ai) for motion trajectories extracted from a gesture image sequence at two hand respectively where x, y are positions with respect to the center of the upper area for each frame, v is velocity and Ai is the area; Each feature vector is then given a numerical label to identify the sign it represents.

We implemented a TDNN, try different preprocessing techniques on the data, trained the network and evaluated, the sum square error is calculated according to difference between the target and the output of the neural networks and used as the performance criteria for ArSL recognition system Training of TDNN is performed on the amount of 50% of the extracted solid trajectories from each gesture using an error Backpropagation Algorithm in conjunction with Levenberg-Marquardt (LM) optimization. The rest 50% of the trajectories are then used for testing. Based on the experiments with 40 ArSL gestures, the output from recognition of training trajectories at each category is shown in Table 3. The total average recognition rate on the training trajectories is 100% and the total average recognition rate on the unseen test trajectories is 70.0%.

CONCLUSION

Our research proposed the use of Time Delay Neural Networks (TDNN) as a classification algorithm to achieve a high recognition rate. Our proposed ideas regarding using certain features and using colored gloves aided the recognition process. For comparison purposes, we compare six categories of Arabic sign language. In our research, we use 50% of our own dataset to train our TDNN Algorithm and 50% for testing. Promising results were obtained using the

approach presented by this research where the recognition rate were 100% at training phase and 70.0% at testing phase.

REFERENCES

AL-Rousan, M., K. Assaleh and A. Tala'a, 2009. Videobased signer-independent Arabic sign language recognition using hidden Markov models. J. Appl. Soft Comput., 9(3): 990-999.

AL-Rousan, M., O. Al-Jarrah and N. Nayef, 2007. Neural networks based recognition system for isolated arabic sign language. Proceeding of International Conference on Information Technology (ICIT, 2007), pp: 345.

Charayaphan, C. and A. Marble, 1992. Image processing system for interpreting motion in American Sign Language. J. Biomed. Eng., 14: 419-425.

Khaled, A., S. Tamer and Z. Mohammed, 2012. Low complexity classification system for glove-based Arabic sign language recognition. Lect. Notes Comput. Sc, 7665: 262-268.

Kramer, J. and L. Leifer, 1988. The talking glove: An expressive and receptive "Verbal" communication aid for the deaf, deaf-blind and Nonvocal. SIGCAPH, 39(Spring): 12-15.

Manar, M., A.Z. Farid, D. Mufleh and A.Z. Raed, 2012. Recognition of Arabic Sign Language (ArSL) using recurrent neural networks. J. Intell. Learn. Syst. Appl., 4: 41-52.

Mohandes, M. and M. Deriche, 2005. Image based Arabic sign language recognition. Proceedings of the 8th International Symposium on Signal Processing and its Applications, pp: 86- 89.

Nadia, R.A. and M.A. Yasser, 2012. Real-Time Arabic Sign Language (ArSL) recognition. Proceeding of the International Conference on Communications and Information Technology (ICCIT-2012), Tunisia, pp: 497-501.

Reyadh, N., H.O. Hussein and J. Shaimaa, 2012. Development of a new Arabic Sign Language recognition using K-nearest neighbor algorithm. J. Emerg. Trends Comput. Inform. Sci., 3(8).

Samir, A. and M. Abul-Ela, 2012. 3D Arabic sign language recognition using linear combination of multiple 2D views. Proceeding of the 8th International Conference on Informatics and Systems (INFOS), pp: 6-13.

Sandjaja, I. and N. Marcos, 2009. Sign language number recognition. Proceeding of the 5th International Joint Conference on INC, IMS and IDC, pp: 1503-1508.

Shanableh, T., K. Assaleh and M. Al-Rousan, 2007. Spatio-temporal feature extraction techniques for isolated gesture recognition in Arabic sign language. IEEE T. Syst. Man Cy. B, 37(3): 641-650.

Tolba, M.F., S. Ahmed and A.E. Magdy, 2012. Arabic sign language continuous sentences recognition using PCNN and graph matching. Neural Comput. Appl. J., 2012: 1-12.