Enhancing The Recognition Of Arabic Sign Language By Using Deep Learning And Leap Motion Controller

Article *in* International Journal of Scientific & Technology Research · April 2020

2 authors:

Ammar Alnahhas
Damascus University

Bassel Alkhatib
AIU

Enhancing The Recognition Of Arabic Sign Language By Using Deep Learning And Leap Motion Controller

Ammar Alnahhas, Bassel Alkhatib, Nazeer Al-Boukaee, Noor Alhakim, Ola Alzabibi, Noor Ajalyakeen

Abstract— Because of the need to build assistive systems for people with special needs, many computer systems have emerged to help understand sign language. In this paper we present an innovative method for recognizing words in Arabic sign language using the Leap Motion device which helps building a 3D model of the human hand using infrared. Our methodology focuses on analyzing the mathematical features derived from the Leap Motion controller, where we take advantage of the chronology of successive frames that can be represented using the extracted features, we process these features using a Recurrent neural network. We present our results which we have performed on real data, the results show superiority of our method Which has achieved outstanding results that outperform previous research. The experiment result shows that the highest average classification rate reached 89% for one-hand gestures, 96% for two hands gestures.

Index Terms— Sign language recognition, Arabic sign language, Deep learning, LSTM, RNN, Leap Motion controller, neural networks..

1 INTRODUCTION

The rate of deaf and hard hearing people in the Arab world is one of the highest in the world, it exceeded 2% of the population in some countries. However, due to lack of awareness and Ignorance of the abilities of those people, the society does not pay enough attention to them, to address this problem, many researchers started to introduce works to support these people by developing useful technologies. Deaf people use sign languages to communicate, a sign language consists of many expressions that can be expressed using actions of the body including hands and face. Sign languages are widely understood by deaf people as it is the only way for them to communicate, but the problem arises when hearing people need to communicate with deaf people, as they rarely understand sign language. Therefore, many researches have emerged in the last decades to support building computer systems that can automatically interpret sign languages and help people understand and communicate with the deaf. These researches used many methodologies varying from image processing to using sensors that detect action of the body.

In this paper we present a novel methodology that can recognize dynamic gestures that represents expression in Arabic sign language using Leap Motion controller in conjunction with deep learning technology, as Leap Motion controller is an infrared sensor that can detect the motion of hands and fingers, and as the motion of the hands is a key element is sign language, using this controller supports recognizing the expressions of this language. To process dynamic gestures, we propose to represent sensory data as a series of frames, each frame is a collection of values that represent features of hand posture in this frame, the resulting series of frames is processed using recurrent neural network, specifically long short-term memory (LSTM) which is a well-known tool to encode time series by extracting latent representation that can be used to classify series into classes which represent expressions of sign language. We build a system that uses our proposed methodology and can recognize expressions of sign language that can be performed by using hands either by one or by two hands. The reset of this paper is organized as follows: the second section shows the previous work in this field in depth, the proposed method is elaborated in section 3 and the results are presented in section 4, finally section 5 concludes the paper.

2 RELATED WORK

In the last years, many researchers were concerned in developing systems that help recognizing sign languages, researches varied in the way they address the problem, they targeted different sign languages, some of researches focused on static gestures where no hand motion is required, while others considered recognizing dynamic gestures where sequence of states is involved. Many researches use Leap motion controller (LMC) to extract features of the hand while others used Microsoft Kinect, and some researchers used images taken from cameras to extract features by using image processing technique. Different classification methods were used to classify feature data, methods varied from neural networks, Bayesian networks and many others. Finally, some researches aimed at recognizing alphabet and number gestures while other targeted gestures that denote words and phrases, which is more difficult to consider as the number of these gestures are not small, table 1 summarizes the researches presented in this study. Researchers in [1] used LMC to recognize American static sign language, they recorded the gestures that represent alphabet and numbers, then classified the data using distance algorithms. researchers in [2] tried to use LMC to recognize static gestures of American sign language as well, but they used recurrent neural network as a classifier. Researchers in [3] uses LMC to build a system that can recognize both manual and finger spelling gestures of American sign language, they used two levels of classification: the first one is a SVM that can detect if a gestures is manual or finger spelling, the second is a RNN that can recognize gestures. Similarly, researchers in [4] used Leap Motion Controller as well, they used both k-nearest neighbor and SVM to classify 26 English alphabet gestures. Researchers in [5] used Leap Motion controller as well, they

- *This paper is prepared in the faculty of information technology engineering – Damascus University – Damascus -Syria with cooperation of he Faculty of Information technology and communication engineering Arab International University - Syria*
- *Ammar Alnahhas: eng.a.alnahhas@gmail.com*
- *Bassel Alkhatib: drbasselalkhatib@gmail.com*
- *Nazeer Al-Boukaee: nazeer250@gmail.com*
- *Noor Alhakim: e.noor.ite@gmail.com*
- *Ola Alzabibi: ola.za9@gmail.com*
- *Noor Ajalyakeen: Noor.ajalyakeen@gmail.com*

built a system to recognize American static gestures and used SVM as a classifier. Researchers in [6] compared three different recognition techniques; geometric template matching, artificial neural network and cross correlation to classify 26 alphabet gestures of American sign language, they reported that geometrical template matching achieved the highest score. Researchers in [7] used a decision tree to recognize gestures of American sign language, each node of the tree is associated with a geometric feature of the gesture, the full tree contained 16 different types of decisions, similarly, researchers in [8] used geometrical matching to recognize static gestures of ASL. Some researchers targeted different sign languages, researchers in [9] presented a system to recognize Spanish sign language using Hidden Markov model, comparing it with KNN and DTW, they showed that HMM achieved good results. Researchers in [10] used LMC to recognize letter signs of Indonesian sign language, they used Bayesian algorithm for classification and achieved good results. While researchers in [11] used a decision tree to classify Thai alphabet signs. Few researchers targeted dynamic gestures recognition, researchers in [12] used LMC to extract features of hand gestures in order to recognize both static and dynamic signs. They used artificial neural network to classify the features, while researchers in [13] worked on a system that uses LMC to recognize dynamic gestures of Indian sign language, the proposed method used KNN as a classification method. Arabic sign language has been addressed in many researches before, Some researchers targeted static gestures of Arabic sign language, authors of [14] developed a part-based hand gesture recognition system using the Leap Motion Controller for Arabic sign language using SVM classifier, they classified static hand gestures of ArSL for letters "alif"-"yah" and digits 0-9. Similarly, Researchers in [15] also used LMC to recognize letters and digit gestures of ArSL, but they compared two different types of classifiers: Naïve Bayes classifier and Multilayer perceptron, they showed that the neural network achieved better results. Some researchers used image processing techniques to recognize Arabic sign language, researchers in [16] proposed to process images of gestures for Arabic alphabet to extract features that can be classified using SVM, similarly researchers in [17] also used image processing to extract features for recognizing ArSL, but they targeted gestures that represent words rather than alphabet, besides, they used HMM as a classifier. Some researchers focused on recognizing dynamic gestures in ArSL, researchers in [18] developed a system to recognize dynamic Arabic sign language using Microsoft Kinect, the recognition process depends on two machine learning algorithms: Decision tree and Bayesian network, besides, they applied Ada-Boosting technique to enhance the recognition of the system. Researchers in [19] proposed a model for both static and dynamic Arabic sign recognition using the LMC. Their system targeted 28 Arabic alphabet, digits from 0 to 10, eight common Arabic signs which are used at the dentist, 20 common nouns and verbs used in the different aspects of life and finally 10 signs which are performed by two hands. They compared the performance of three classifiers: SVM with poly kernel, -K-Nearest Neighbor (KNN) and Multilayer Perceptron. Authors of [20] used dual LMC to recognize 100 different dynamic Arabic signs, they used both Bayesian approach and a Gaussian Mixture Model. As we showed most of the previous researches focused on recognizing static sign language that include alphabet and numbers in different languages and only few of them focused on recognizing dynamic gestures. And to our knowledge, no previous work discussed the recognition of dynamic Arabic sign language that include vocabulary gestures using leap motion controller with recurrent neural network. So, we are going to present a novel methodology that can use features extracted from consequent frames each represent a state of hands in a point of time, and use a LSTM network to process these features to recognize dynamic gestures that they represent.

TABLE 1
SUMMARY OF RELATED WORK PRESENTED IN SECTION 2

#	Language	(S)tatic/(D)ynamic	Classification method	Feature extraction	Word/alphabet
[1]	American	S	Distance	LMC	Alpha
[2]	American	S	RNN	LMC	Alpha
[3]	American	S	SVM	LMC	Alpha
[9]	Spanish	S	HMM	LMC	Alpha
[5]	American	S	SVM	LMC + sensors	Alpha
[10]	Indonesian	S	NB	LMC	Alpha
[12]	American	D+S	ANN	LMC	Alpha
[6]	American	S	GTM	LMC	Alpha
[7]	American	S	Decision tree	LMC	Alpha
[4]	American	S	SVM	LMC	Alpha
[11]	Thai	S	Decision tree	LMC	Alpha
[8]	American	S	Distance	LMC	Alpha
[13]	Indian	D	KNN	LMC	Alpha
[14]	Arabic	S	SVM	LMC	Alpha
[15]	Arabic	S	NB	LMC	Alpha
[19]	Arabic	D	SVN + KNN + ANN	LMC	Alpha + words
[20]	Arabic	D	NB + GM	LMC dual	Words
[16]	Arabic	S	SVM	Image recognition	Alpha
[17]	Arabic	S	HMM	Image	Words
[18]	Arabic	D	DT + NB	Kinect	Words

3 PROPOSED METHODOLOGY

Our proposed methodology of recognizing dynamic Arabic sign language is based on recording hand motion using Leap motion controller, the recorded motion is converted into sequence of frames, each frame is a state of the hand in a point of time, so the sequence of frames represents the motion of the hand during the time of the dynamic gesture. Each frame is represented as a vector of geometrical values that represent hand posture, that is, the position of certain part of the hand along with the value of angles between parts of the arm, as will be viewed later in this paper. The sequence of vectors that represent the gesture is passed through a LSTM network that is known to be suitable for classifying sequence of temporal data. We present the problem of recognizing gestures as a classification problem, where each gesture is represented as a class, so the number of classes is equal to the number of gestures that the network is able to recognize. The output of the network determines the gesture that match the hand motion being recognized. Our methodology is tested against gestures that need one hand motion and other gestures that need two hands motion, it proves efficiency in both cases as we will discuss in the testing section. Figure 1 shows the flow of proposed methodology.

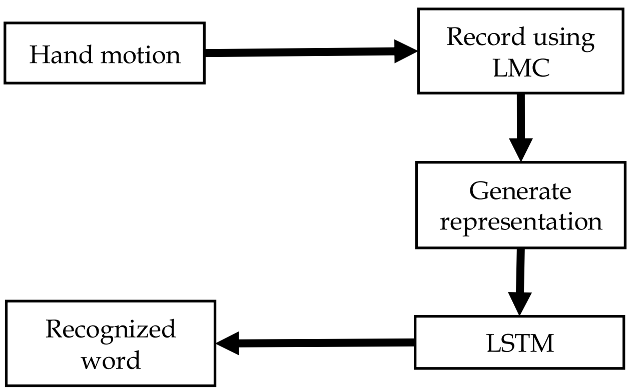

Fig. 1. The proposed methodology

3.1 Detecting gestures

The first step of our methodology is to capture the posture of the hand in each point of time to record the dynamic gesture. To do this we use Leap motion controller, which is a device that connects to a computer with USB connection, it is a tool for hand movement detection, it can read hand's skeleton with 2D or 3D coordinates within a space, thus defining joints and bones positions during a particular movement. LMC is designed basically to be placed on a desktop to recognize hands above it, nevertheless, it can be used in a mode called VR mode, where the device is mounted above a glasses frame, we use tis mode in our experiment as we noticed that it is more suitable to detect hand movement of sign language in this mode. Figure 2 shows the recognition device we use.

Fig. 2. LMC in VR mode

3.2 Generating gesture representation

LMC can detect the skeleton of the hand, it can find the position of each joint in a defined coordinate system, this system uses the right hand model with Euclidean coordinates, The principle of coordinates is at the top of the controller, Both x-axis and z-axis are located in the horizontal plane where the x-axis is parallel to the long edge of the device, while the y-axis is vertical to them and the z-axis increases its positive values towards the user as shown in figure 3.

Figure 4 shows the points that LMC can detect in the skeleton of the hand, along with elbow position, these positions can represent the hand in a point of time, which we call it a frame. In each frame we record the following values:

- el_t: The position of the elbow at time t.
- wr_t: the position of the wrist at time t.
- pa_t: the position of the palm at time t.
- ar_t: arm direction at time t, which is wr_t-el_t
- pd_t: palm direction at time t, which is pa_t-wr_t
- p_t,r_t,y_t: pitch, yaw and roll angles of the palm at time t.
- $b1_t$,$b2_t$,…,$b14_t$: angle value between bones of the fingers, two for the thumb and three for each other finger, they can be found using positions of the joints taken from LMC.

Formally, the frame at time t denoted f_t is a vector of 32 values mentioned in the former list. Note that positions and directions are vectors with three elements and angles. Dynamic gesture is a hand movement that last for a period of time, so, it can be represented as a sequence of frames, each represent the state of the hand in a time point, in our methodology we suppose that all gestures take the same period of time, actually this is not true, but it is necessary to have the same sequence length when using LSTM networks as we will discuss later. Besides as we are detecting isolated gestures in our study, it is easy to use the last real frame of the gesture as a padding to extend the length of the sequence to the intended length, and this is done automatically as user will usually keep hand in the last position of the movement for a period of time. A gesture G that is performed in a fixed period of time T, can be represented as a discrete sequence of states. We discretize the period T into fixed points of time t1, t2, …., tn where:

$$t_n - t_{n-1} = \frac{T}{n}$$

Now G can be represented as a sequence of frames each represent hand posture in a point of time:

$$G = (f_{t_1}, f_{t_2}, \dots, f_{t_n})$$

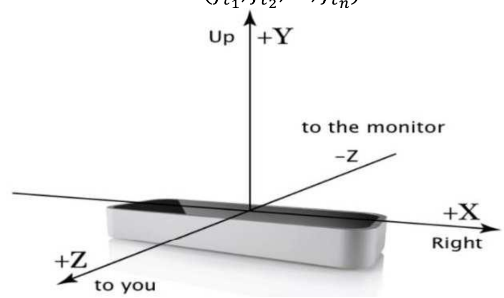

Fig 3 [21], Coordinate system of Leap Motion Controller

Fig 4 [21], Leap Motion hand tracking

3.3 Classification
To recognize dynamic gestures, we propose to view the problem as a multiclass classification task, each gesture is represented as a class, we use a long short term memory (LSTM) which is a type of Recurrent neural networks (RNN) as a classifier, LSTM is known to be suitable classifier for temporal data, it can determine the relations between different type steps and find internal representation of the whole series which can be used as input to an artificial neural network that can classify this representation and find the suitable class. The structure of our classifier is shown in figure 6, the first layer of the network is a LSTM layer. The input of this layer is a 32□20 matrix, where 32 is the number of features in each time step, that is he number of features in each frame, and 20 is the number of frames which is the number of time-points we choose to discretize the period of gestures into. The number of units of LSTM layer is 32 which is the length of the output of this layer. The next layer is a fully connected layer that takes the output of the LSTM as input, the length of the output is the number of gestures to recognize in the dataset and is different in each experiment as shown in the testing section. The activation function of the second layer is SoftMax which is suitable for multiclass classification tasks.
During the training phase of the network, we use gesture representation as input and a one hot vector which represent the known gesture as a target. The network is trained to minimize the cross-entropy error.

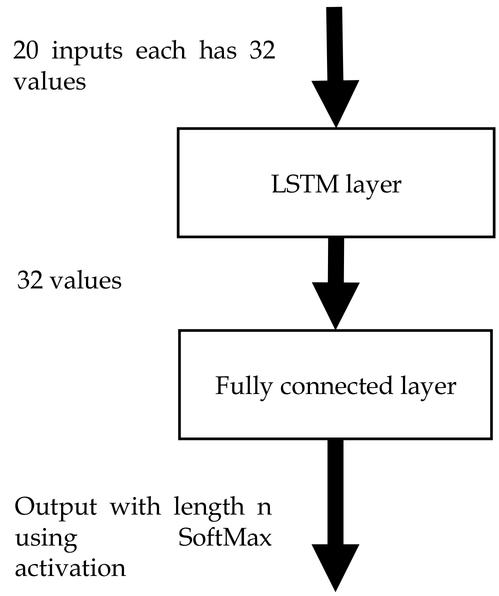

Fig. 5. Structure of classification neural network

4 Testing and Results
To test the proposed methodology, we conduct an experiment with real data, we constructed a training dataset and evaluate the results on a testing dataset.

4.1 Dataset
Our dataset consists of 44 signs that are commonly used in Arabic sign language, 29 of them are one hand gestures, and 15 are two hand gestures. We used two different networks for one hand and two hand gestures, because the second one has an input with double length as the first one, that is, each frame is represented by features of both hands. We record 10 samples for each gesture by 5 different signers to check that the model can correctly classify the gestures of any user. The dataset is split into two parts 80% for training and 20% for testing.

4.2 Evaluation metrics
Because we record equal number of samples for each gesture, we can use the accuracy of the system to measure the performance. The accuracy is the ratio of the test cases where the gesture proposed by the model is equal to the target gesture of the sample, formally, if Sto is a sample from the test dataset TDS where its target gesture is t and its output is o, then the accuracy is:

$$acc = \frac{\sum_{g \in G} |\{s_g^g\}|}{|TDS|}$$

4.3 Results
Table 2 shows the results of the methodology along with dataset details, we show that the accuracy of our methodology for one hand gestures is 89% where it achieves 96% for two hand gestures.

TABLE 2
RESULTS OF PROPOSED METHODOLOGY

	One hand	Two hands
Number of vocabularies	29	15
Number of samples	290	150
Number of training samples	232	120
Number of testing sample	58	30
The accuracy for training data	100%	100%
The accuracy for testing data	94%	96%

Besides we show the precision of our methodology for some vocabulary in both one hand and two hand gestures in table 3, the individual precision values show that many gestures achieved good results, which indicate that the proposed methodology is scalable and will work for other gestures not in the experiment.

TABLE 2
Precision for individual vocabulary of one hand gestures using our system

Arabic vocabulary	Translation to English	Precision
سأل	Asked	100%
أمسك	Catch	100%
عَلا	Raise	75%
بحر	Sea	100%
نثَر	Seed	100%
أخذ	Take	100%
سافر	Travel	100%
أنتم	All of you	100%
الجنوب	south	100%
غداً	tomorrow	100%
تفضل	welcome	100%
الغرب	west	100%
البارحة	yesterday	100%

We implemented the methodology in a real system that can help recognizing Arabic sign language, figure 7 shows some gestures when detected by our system.

TABLE 3
Precision for individual vocabulary of two hand gestures using our system

Arabic vocabulary	Translation to English	Precision
أضاف	Added	100%
وازن	Balance	100%
صعوبة	Difficulty	100%
هدف	Goal	100%
أزاح	Push	100%
نَشَرَ	Spread	100%
كتب	write	75%

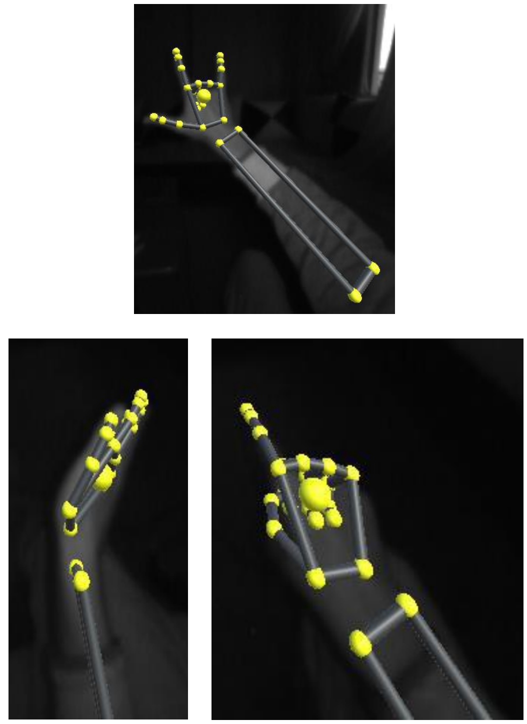

Fig. 6. Samples of detected gestures by our system

4.4 comparing results with previous work

To evaluate the efficiency of our system, we compare the results with previous researches that are similar in their goal, which is recognizing dynamic gestures that represents words and phrases in Arabic sign language, we compare the results with three different researches:

- [19] in which researchers used LMC and SVM as a classifier.
- [20] where researchers used Gaussian mixture as a classification method.
- [18] where researchers used Microsoft Kinect to detect hand motion and a Bayesian network as a classifier.

Figure 7 shows the comparison result, we can notice that out method outperformed most of the other researches.

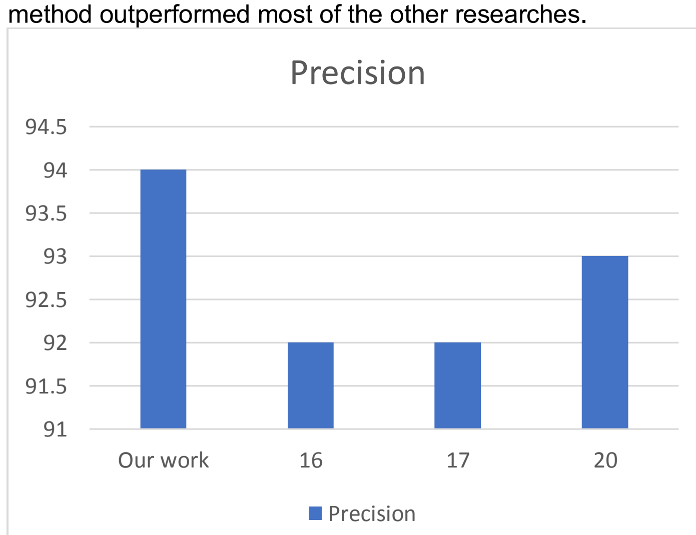

Fig. 7, comparison with related work

5 Conclusion

It is important to build system that can recognize sign languages to help deaf people, we showed in this work a novel methodology to recognize dynamic gestures of Arabic sign language, the methodology is based on detecting hands elements using Leap motion controller. The gesture is represented as a sequence of frames to reflect its temporal nature. We used a neural network classifier that depends on LSTM layer to encode the sequence and find the matching gesture by doing a multiclass classification. The Results we presented in this paper reflects the efficiency of the proposed methodology. And showed that our methodology can accurately recognize dynamic gesture of ArSL, future research may be conducted to develop the methodology so that we can recognize overlapped gestures and build complete sentences out of sign language gestures.

6 References

[1] Auti, A., et al., Sign-Talk: Hand gesture recognition system. International Journal of Computer Applications, 2017. **160**(9).

[2] Avola, D., et al., Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Transactions on Multimedia, 2018. **21**(1): p. 234-245.

[3] Kumar, P., et al. Real-time recognition of sign language gestures and air-writing using leap motion. in 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). 2017. IEEE.

[4] Chuan, C.-H., E. Regina, and C. Guardino. American sign language recognition using leap motion sensor. in 2014 13th International Conference on Machine Learning and Applications. 2014. IEEE.

[5] Quesada, L., G. López, and L. Guerrero, Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments. Journal of Ambient Intelligence and Humanized Computing, 2017. **8**(4): p. 625-635.

[6] Khan, F.R., H.F. Ong, and N. Bahar, A sign language to text converter using leap motion. International Journal on Advanced Science, Engineering and Information Technology, 2016. **6**(6): p. 1089-1095.

[7] Funasaka, M., et al. Sign language recognition using leap motion controller. in Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). 2015. The Steering Committee of The World Congress in Computer Science, Computer ….

[8] McNeil Jr, S.H., Sign Language Static Gesture Recognition Using Leap Motion. no. November, 2017.

[9] Parcheta, Z. and C.-D. Martínez-Hinarejos. Sign language gesture recognition using HMM. in Iberian Conference on Pattern Recognition and Image Analysis. 2017. Springer.

[10] Wibowo, M.D., I. Nurtanio, and A.A. Ilham. Indonesian sign language recognition using leap motion controller. in 2017 11th International Conference on Information & Communication Technology and System (ICTS). 2017. IEEE.

[11] Tumsri, J. and W. Kimpan. Thai sign language translation using leap motion controller. in Proceedings of The International MultiConference of Engineers and Computer Scientists. 2017.

[12] BIRRY, R.A.K.A., ASL RECOGNITION USING LEAP MOTION AND HAND TRACKING MECHANISM. International Journal of Advances in Electronics and Computer Science, 2017.

[13] Sonawane, T., et al., Sign language recognition using leap motion controller. Internation J. Advance Res. Innov. Ideas Edu., 2017. **3**(2): p. 1878-1883.

[14] Khelil, B., et al. Hand gesture recognition using leap motion controller for recognition of arabic sign language. in 3rd International conference ACECS. 2016.

[15] Mohandes, M., S. Aliyu, and M. Deriche. Arabic sign language recognition using the leap motion controller. in 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE). 2014. IEEE.

[16] Alzohairi, R., et al., Image based Arabic Sign Language Recognition System. International Journal of Advanced Computer Science and Applications (IJACSA), 2018. **9**(3).

[17] Assaleh, K., et al., Continuous Arabic sign language recognition in user dependent mode. 2010.

[18] Hisham, B. and A. Hamouda, Supervised learning classifiers for Arabic gestures recognition using Kinect V2. SN Applied Sciences, 2019. **1**(7): p. 768.

[19] Hisham, B. and A. Hamouda, Arabic Static and Dynamic Gestures Recognition Using Leap Motion. JCS, 2017. **13**(8): p. 337-354.

[20] Deriche, M., S.O. Aliyu, and M. Mohandes, An Intelligent Arabic Sign Language Recognition System Using a Pair of LMCs With GMM Based Classification. IEEE Sensors Journal, 2019. **19**(18): p. 8067-8078.

[21] developer, L.M. Leap Motion coordinate system. [cited 2020 2/8/2020]; Available from: https://developer-archive.leapmotion.com/documentation/csharp/devguide/Leap_Coordinate_Mapping.html.