ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse37952020.pdf
https://doi.org/10.30534/ijatcse/2020/37952020

Automatic system for Arabic Sign Language Recognition and translation to spoken One

Mokhtar M. Mohamed
Assistant Professor, Department of Natural and Applied Sciences, Unaizah Community College, Qassim University, Saudi Arabia, h.hassien@qu.edu.sa
Assistant Professor Department of Computer and Information Systems, Sadat Academy, Assiut, Egypt

ABSTRACT

The deaf-mutes person suffers from inability to speak as well as inability to hear others. This disability creates a barrier that prevents the disabled from integrating with society. Therefore, the owners of this disability attempt to break this barrier by using many methods to try to communicate with others. The most famous of these methods is the sign language which is relying on the body language. The disabled person uses facial expressions and hand gestures to express his needs. In this project, we will try to build a computerized system depending on the depth-measuring cameras and computer vision techniques to capturing and segmenting the pictures of the facial expressions and hand gestures. The segmented gestures are classified and stored in an electronic library after they have been recognized and linked with their corresponding spoken and written Arabic words. The computerized system will be learned with gestures and its corresponding spoken and written Arabic words. A gesture dataset of 40-segmented words that frequently used in the life by the hearing-impaired person was utilized for testing the system. We get in our consideration that each word may has a different occlusion state. The experimental results illustrated that the proposed system has a recognition accuracy over 90%

Key words: Sign language, Segmentation, Tracking, Recognition, Kinect, Fuzzy C-mean, Neural Networks

1. INTRODUCTION

Sign Language (SL) is not like spoken languages. It depends on vision instead of hearing. Deaf person uses SL to communicate with deaf community or with hearing one. The popular belief that SL is an international language is not the truth. This adds further difficulties on either deaf community or hearing one. One can suggest written communication instead of SL to overcome its difficulties. In fact, this mean of communication is very cumbersome and most of deaf people unknown writing spoken language. In addition, written communication is very slow in real time conversation. Signer can communicate with other communities through SL by moving their hands, moving their body and by facial expressions. Moving hand or body produces gestures, which can be translated to a spoken language. Actually, hearing people who understand SL are very rare. Only a skilled person can do this translation. Recently, many innovations tried to translate automatically from SL to spoken language and vice versa utilizing computer vision techniques. These techniques are used in somehow to construct a computerized system for SL recognition. This system assists the deaf person to communicate easily with hearing one. SL recognition systems can be classified to two different categories. The first one is for a single gesture while the second one is for a sentence of gestures.

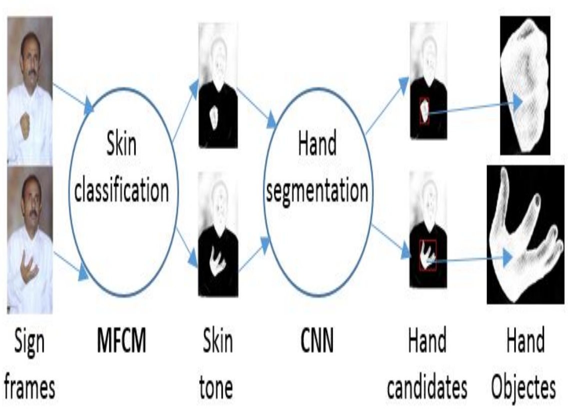
In both previous categories, the sign gesture can be composed from hand motion in addition to hand shape. This type of sign is called manual components. On the other hand, the sign gesture which composed from hand motion, hand shape in addition to facial expressions is called non manual components. In this work, we study the isolated sign gesture using manual components for Arabic sign language (ASL).

The SL recognition system based on manual components processes the sign gesture according to hand movement simultaneously with hand shape. The process tracks the sign gesture from the starting to stop of the sign as illustrated in [40]. Our proposed system for sign gesture recognition based on isolated sign for ASL is shown in Figure 1.

For each sign gesture, the proposed system tracks the hand in each frame. The tracking mechanism uses two different models: one is the observation model and the other is the dynamic model. The observation model contains information about the hand motion and pre-trained hand samples by Convolutional Neural Networks (CNN). While, the dynamic model contains information about hand location in the interest frame and its velocity. This mechanism helps us to determine the interested region based on motion information. On other hand, it confirms the presence of hand in the interested region based on the trained CCN.

The tracking mechanism begins its job by segmenting the hands from the backgrounds. Each segmented hand is assumed as a tracking candidate state. This candidate contains information about hand location. According to this information, the hand patch images is extracted from the entire patch frame. The extracted hand patch images will be assumed as the input of the observation model. In this stage of tracking process, the pre-trained CCN will produce a hand feature vector (HFV) for highest likelihood hand patch. Each HFV contains information about hand shape and location for certain hand patch. Finally, the detected HFV is used to extract the corresponding audio translation from the pre-build database.

The overall proposed system is dedicated to Arabic deaf people as mentioned before. Therefore, we consider the Arabic sign language (ASL) vocabularies as depicted in dictionary [19]. The system is equipped with depth camera and monitor. It may be placed in lecture halls, conferences and airports. The system can be impeded in mobile devices to help disabled person to communicate at home and market as well as hold chat conversations over the internet with hearing one in easy manner.



The rest of the paper is organized as follows: state of arte and recent related works are discussed in section 2. A description for the used CCNs is illustrated in section 3. The Proposed hand tracking system and output is explained in section 4. Experimental results, discussion and database used in this work is presented in section 5. Finally, our conclusion for the proposed system is given in section 6.

2. PREVIOUS WORKS

A computerized based recognition process for the grammatical structures of Arabic Sign Language (ASL) is a challenging problem. Early, researchers have concentrated their attention to recognize the finger gestures, facial expressions, isolated gesture, and words of interrogative like How, Where, When, and What. Also, another challenge is verification of sentences and its grammatical structures. In addition, some classifiers tended to surrogate the hand gestures for the deaf-mutes people and the objects classes information about its location, its movement and its appearance.

In general, gathering information about sign gesture can be achieved through two different methods such as direct measurement or vision base. Direct measurement depends on extra equipment like motion capturing sensors and data gloves. While vision based detects and tracks sign gesture in natural environment without extra equipment.

Through direct measurement methods, obtaining spatial information about finger movement, hand shape and body posture is accurate. Although researchers who use this type of gathering data are easily obtain information about finger, hand and body posture in accurate manner. They found it inapplicable in natural circumstances. Because deaf person will face several difficulties to wear devices, which restrict his movement in addition to its complicated to setup [20], [21], [22], [23].

Due to the several difficulties in applying direct measurement methodologies, researchers preferred to use vision based methodologies. In this type of sign gesture recognition, researchers depends on a camera or a group of cameras to capture images and store it in a sequence of two-dimensional images. So, it easy to capture gestures and record as simple as we do in our real live. However, vision base methods have challenges such as camera quality, noise, illumination. Interested region and tracking. Most of researches based on vision use color images as input to their systems. The need for color image is to facilitate classification of hand candidate and face expressions based on the human skin-tone color [24], [25], [43], [46]. Other researches use colored gloves and face masks to achieve sign gesture classification process [26], [27], [28], [44], [45].

Kelly et al. [29] proposed a system for SL recognition. They used the mean shift algorithm in tracking colored gloves. They segment the glove region by using the back projection technique, which computed using the mean shift algorithm. From the segmented region, the hand contour can be extracted. Oya et al. [1] explained another similar work to track the hand gestures. The authors used gloves with blue and yellow colors. They utilized the histogram bins and a predefined threshold to segment the gestures from the image. By this approach, they utilized colored glove to find out the hand and finger positions.

Zou et al. [30] used mean shift with Kalman filter to track hand. They used mean shift algorithm in updating the sign gesture model Whereas, Kalman filter is used in tracking hand. They used mean shift algorithm due to its computational efficient and its robustness against pose variance. On the other hand, it failed to track with efficient object that has extreme illumination variance and that has fast moving. The disadvantage of the mean shift is explained in [31]. The authors returned this disadvantage to the dependence of mean shift algorithm on histogram information and ignore spatial information. Nadgeri et al. [32] used CAMShift algorithm to overcome the illumination changes. Also, CAMShift use adaptive window with not fixed size. The window size is changed according the object size.

A system based on vision was proposed in [2] to segment gestures based on skin color information only. They used a Kalman filter [3] technique to track the gestures. Kalman filter is one of the earliest algorithms that still used in visual tracking. It estimates the linear states, dynamic system in discreet-time in efficient manner. Kian et al. [33] explained a real-time system. They used Kalman filter to track linearly the hand's location based on skin color. As mentioned in [34] kalman filter has a strong robustness and advantages under occlusion states. However, according to [35], kalman filter and its variants have uncertainty in case of Gaussian distribution and cannot maintain multiple hypotheses.Another system in [4] was proposed depending on snake methodology to separate and tracking the face expressions and hands gestures. The snake tracking technique has an advantage in solving the occlusion problem. A system in [5] combines the skin segmentation algorithms with the frame difference. This system uses a probabilistic method to predicate position of the face and hands.

Another filter is call Particle filter has been applied in several tracking system [36], [37], [38], [39]. The authors of these researches mentioned that particle filter has strong performance in case of non-linear dynamic and non Gaussian noise. Particle filter performance is based on two models, dynamic model and observational model. Dynamic model produces a distribution model for the noise. While, observational model produces a likelihood between target object and particle object. The object location is obtained according to the particle with high weight. A research for real time hand tracking is proposed in [40]. The authors used mean shift with particle filter as an estimator for non linear posterior density. The research in [1] used a joint particle filter depends on two ratios for calculation the likelihood model of hand and head. They used the mean shift to obtain a minimum number of particles. In [41], authors proposed a tracking technique depends on gravity optimized particle filter.

Microsoft Kinect camera was used in [4] to extract hand features based on its appearance and to track hand position in both 2D or 3D. the classification process is achieved by hold a comparison between hidden Markov model and the sequential pattern boosting. Another research used Microsoft Kinect in [2]. The authors built a system for Chinese sign language recognition. They construct Chinese sentences by using the language models with 3D trajectory movement of the hand. The trajectory is registered and matched with the stored in a gallery.

The depth information also was used in [6] to simplify the problem of segmentation and tracking the head and hands. The hands gestures are segmented according to the idea that the hands would be near the camera. Another system in [7] used the depth camera to recognize the sign language words that are constructed with the finger spelling. The Kinect camera was used in [8] to enhance the tracking process. The authors used skin color information as the main feature for segmentation. They claim that the results are improved on the used dataset, which consists of many different complex gestures. A mobile-based system for interpreting between hearing and deaf-mutes people was proposed in [9]. The system is beneficial for achieving in communication on-the-go through the real time mode. Due to the processing limitation of the mobile devices, the authors connected the system to a cloud-based framework. They utilized the cloud system powerful resources by delegating all the heavy work of the proposed system.

The authors in [10] represents an automatic visual sing language recognition system that translate segmented Arabic sign words into text. Their system was constructed from four main stages: hand classification, feature extraction, tracking and segmentation. Based on the face color tone for hand, they used a dynamic technique for skin detector segmentation. They identified and tracked the hands according to a skin-blob technique. A system to translate from sign language to text and vice versa was proposed in [11]. This system depends on Canny's edge detection algorithm [12] to segment the hand gestures. It also built a library of gestures and the corresponding text.

The authors in [13] proposed a system for converting the finger gesture spelled word to speech style and vice versa. The proposed system gets the hand gesture in continuous image stream and converts it to letter of alphabets. The system used a tracking algorithm based on continuously Adaptive Meanshift (Camshift) [14] to jointly track the hands and face at same time. The authors [15] developed an Arabic Language recognition system that based on a modern digital sensor called Leap Motion. It concentrates on the general issues in vision-based systems such as lighting and skin color. The Leap motion sensor captures hands and fingers motions in a 3D digital format. It collects the 3D digital information in individual frame of motion.

Spatial and temporal features are passed into a Multi-layer perceptron Neural Network (MLP). The system cannot track non-manual features. To do this it needs more sensors to track facial expressions and body poses as a non-manual features. Deep learning models [16] used deep layers architecture with the iterative nature. They used CNN to recognize the fingerspelling on American Sign Language dataset. They used a dataset of 31 alphabets and numbers. The system was fed with depth maps for the data which was captured by depth sensor and digital camera. The system achieved 83.5% to 85.5% accuracy. An Italian Sign Language recognition system [17] was proposed a method using CNN. It used a dataset consists of 20 Italian gestures which were recorded by Microsoft Kinect. The system consisted of two CNNs; one for hand features extraction and another for upper-body features extraction. The results then were concatenated to enter a classical Neural Network Classifier.

2.1 Modern SL Sensors

Several works in SL recognition systems around the world utilized different sensors to deal with the input signs.Through last decade, due to the rapid growing in technology, many modifications are achieved to boost the accuracy of the sensors through the signs features capturing process. Examples of these sensors are Leap motion and Kinect sensors those capture the hands and body poses. Figure 2 illustrates a sample of Kinect image and another of Leap.

Figure 2: Kinect versus Leap images. a) Image of Kinect. b) Image of Leap.

3. CONVOLUTIONAL NEURAL NETWORK (CNN)

The name of convolution neural network comes from the name of the mathematical convolution operation. CNN is designed to using in image classification and recognition. CNN is mainly constructed from three main layers: convolution layer, pooling layer and connected layer. The famous features are edges and textures. It uses a kernel or filter in features extracting operation by applying convolution operation on the input image.

Lionel et al. [42] mentioned that, CCN is inspired by the visual cortex of human brain. In CCN, the artificial neurons will connect to a local region of the visual area, called a receptive area. This is achieved by performing discrete convolutions on picture with values as trainable weights. Several filters are applied for each channel and with activation functions will form the feature map. The convolution process is followed by a pooling process. Pooling process pool only the interesting information of the feature maps. Therefore, CCN is a feature extraction model by deep learning algorithms. It very successful in case of object recognition. Many applications use CCNs as a core to achieve its job such as Google, Facebook and twitter.

In general, kernel is a two-dimensional array of values that will be considered as weights. The convolution operation is performed by sliding the kernel over the image as shown in Figure 3. The convolution layer output is a map of features. Each region under sliding and convolution process is called interested region (IR). The convolution process is achieved by using the following formula.

$$Z_{ij} = (I * K)_{ij} = \sum_m \sum_n I_{i-m,j-n} K_{mn} \quad \text{-1}$$

where I is the input image, and K is the kernel. The output of each layer in CNN can be expressed as following:

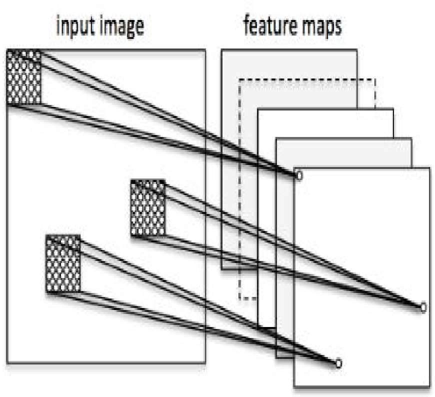
$$y_i^l = f(z_i^l) \quad \text{where} \quad z_i^l = \sum_{j=1}^{l-1} w_i^j x_j^{l-1} \quad \text{-2}$$

Where y is the layer output, z is the activation function, i is neuron in the layer l, w is the weight and x is the input data.

$$w = \{w_{ij}^{l} : l = 1,3,...L-1;\ i = 0,1,...I;\ j = 0,1,...,J\} \quad \text{-3}$$

$$x = \{x_{j}^{l} : l = 1,3,...L-1;\ j = 0,1,...J\} \text{-4}$$

The second layer in the CCN is the pooling layer. The main purpose of this layer is to simplify the feature map coming from the convolution layer. It uses the max, sum or the average operation in emphasizing the feature. The third layer is the fully connected layer. The main purpose of this layer is to convert the two-dimensional feature map to one-dimensional. This form is suitable for pre-learned feature classification decision.



In our work, we used the dictionary for Arabic sign language [19]. We used the some gestures from dictionary as a ground truth in training stage of hand sign recognition. The dictionary is released as groups of gestures' pictures. Each group of pictures represents a class of social situation. We select more than 40 gestures achieved by one hand and more than 10 gestures achieved by both hands. We use this dataset in training CCN. The system was tested using a real gestures achieved by our colleagues.

4. PROPOSED METHOD

In this section, we utilized the Kinect technology. The camera will capture frames for the gestures of deaf-mute person. Figure 4 show the sign Arabic Language alphabets le tters gestures; Figure 5 shows a sample of Arabic word gestures. The gestures images were processed by computer vision segmentation and classification techniques. We modified the Fuzzy C-mean algorithm (FCM) for classification by incorporating prior information about the SL gestures shapes for Arabic language. As a preprocess stage.

The proposed system was learned by prior information for the sign language gestures and its translation to the Arabic words. The learning system depends on three hidden layers of Conventional Neural Network (CNN). We built an electronic library contains sign language gestures and its corresponding Arabic words in two different styles, written and audio. After the system recognize the gesture, it retrieves its translation from the library. The proposed system able to translate from Arabic audio word to sign language by recognizing the word and bringing its corresponding gesture from the library. The system starts when the Kinect camera captures hands gesture. The Camera produces a real time stream about body motion, object depth information, and skeletal motions.

A computer vison framework was developed to detect hands and fingers. It segments and tracks them within its field of scope. The framework achieves movement-tracking information as a stream of frames. The tracked data frame contains the sign orientations, measured positions and other information about each object detected in present frame. The detected hand and finger are represented in a single frame. We can brief our strategy by the following steps on still images.

Step-1: Capturing frames of gestures.

Step-2: Image denoising to refine the image appearance.

Step-3: Classification human skin using FCM

Step-4: Segmenting Signer face using CNN

Step-5: Segmenting hand gesture using CNN

Step-6: Classifying gestures using CNN.

Step-7: Recognizing gestures by comparing with items in the pre-built dataset.

Step-8: Retrieving the corresponding translation for the captured gesture.

4.1 Gesture Classification Based on FCM

In this system, we modify the standard FCM algorithm objective function to incorporate a priori information about the human skin color tone. The new objective function of the FCM has the following form:

$$j = \sum_{c=1}^{Nc} \sum_{i=1}^{N} w(\phi_c)\, u_{ic}^{m} \cdot d_{ic} \text{ -5}$$

where $d_{ic} = \|y_i - v_c\|^2$ and $w(\phi_c)$ is a term that is function of representing a priori information on human skin tone class's.

center v_c and u_{ic}^m is the fuzzy membership with tuning fuzziness m for the pixel i, y_i is the data pixel value. N the number of frame pixel. Finally, Nc is the number of classes. The constraint on the membership u is typical the sum overall classes equal one:

$$\sum_{c=1}^{Nc} u_{ic} = 1 \qquad -6$$

As we illustrated in [18], we assume that the typical and often logical Gaussian distribution for the class centers. So W takes the form of the reciprocal Gaussian probability density function for the class's center as follows:

$$w(\phi_c) = \sigma_c \sqrt{2\pi} e \frac{(v_c - u_c)^2}{2\sigma_c^2} \qquad -7$$

where $w(\phi_c) = \{u_c, \sigma_c\}$ is the Gaussian parameters: u_c is the mean value of the class centers and σ_c is their standard deviation.

Figure 4: Example for Arabic SL letters

$$u_{ic} = \frac{\dfrac{1}{w(\omega_c) d_{ic}}}{\sum_{c=1}^{Nc} \dfrac{1}{w(\omega_c) d_{ic}}} \qquad -8$$

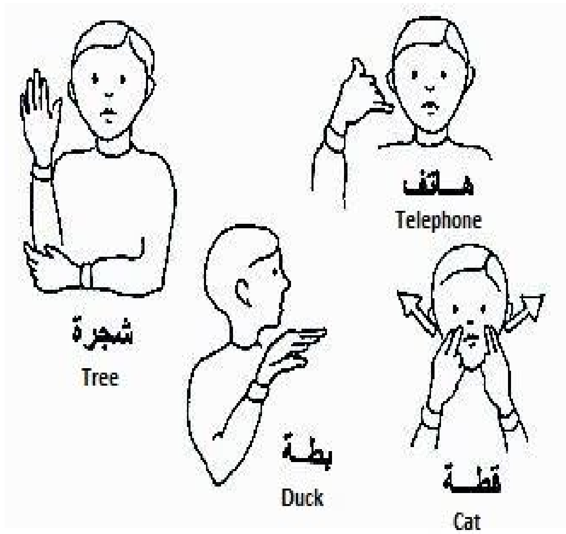


Figure 5: Examples for Arabic SL words

5. RESULTS

In this paper, we developed a fuzzy framework for segmenting, classifying and recognizing the Arabic deaf gestures. A pre-learned gesture database was constructed using machine learning algorithm depending on conventional neural network (CNN). The Arabic gestures are automatically detected and recognized. We built and evaluated our proposed framework and its performance on a machine with specifications, Core i3, 4 GB RAM, Windows 10 with 64 bit, and Matlab 2015. Our database contains 20 images for Arabic SL words and images for all Arabic SL letters. Half of the gestures in our database was used in training stage and the other half was used in testing stage.

We ran several experiments to reach the best performance. Through each experiment, we tried to tune the CCN parameters such as learning rate, decay and momentum. The input images have different depths and different occlusions. Unfortunately, there are not any standard database for Arabic SL gestures. Therefore, we could not compare our proposed performance with state of arts. Some of our developed system results on Arabic words are listed table 1 and some selected letters results are listed in table 2.

In table 1, we can note the low performance of recognition rate of the words duck and cat. We attributed this to the complex of gestures of duck and cat. While, duck gesture needs to adjust the capturing angle to emphasize the fingers details, the cat gesture needs more than one frame to give details about hand movement. In table 2, we note that recognition rate of the letters (ج) and (خ) is less than rate of letter (ح). This is attributed to the small details in outer and insider edges. These small details are canceled through any morphological operation. The same situation is for the couple letters (د) and (ذ), the couple letters (ص) and (ض), and couple letters (ف) and (ق).

Table 1: Samples of Arabic SL words

Arabic word	English meaning	Results
شجرة	Tree	94.37
هاتف	Telephone	88.61
بطة	Duck	76.42
قطة	Cat	68.83

Table 2: Samples of Arabic SL letters

Arabic Characters	gestures	Results
أ		95.21
ب		95.74
ج		84.71
ح		86.35
خ		84.22
د		87.42
ذ		86.33
ص		88.23
ض		86.06
ف		85.45
ق		87.14

6. CONCLUSION AND FUTURE WORKS

In this research, an efficient system for Arabic Sign Language Recognition with Kinect information was developed. The system uses the shape and depth information for the gesture as input. The architecture extracts spatial-temporal features from the input. The gestures are classified and segmented from the background. Then, the hands are tracking in pre-defined number of frames. The gesture is recognized by matching with the corresponding translation in the dataset. Finally, the retrieving process achieves the conversion between gesture and its corresponding spoken word. The system achieves an accuracy rate greater than 90%. This accuracy can be boosted with more training samples added to the dataset to overcome the variety of the signer and the environment surrounded. The lack of standard datasets or a benchmark dataset on SL for Arabic language makes it difficult to compare our results with other works. As a future work, we tend to incorporate several vocabularies. In addition, hand tracking and segmentation canbe enhanced to overcome occlusion of the hand. Information about other body parts. such as the arms and more facial expressions can be included to boosrecognition accuracy. Finally, we tend to achieve a real-time prototype version.

ACKNOWLEDGEMENT

The author gratefully acknowledges Qassim University, represented by the Deanship of Scientific Research on the material support for this research under the number (3769) during the academic year 1439 AH/ 2018 AD

REFERENCES

1. Oya. A, Isamail A. Ari, LaleA. ,Bulant S., Alexandre B. Alice C., Pavel C. Ana H., Francios F., "**SignTutor: An Interactive System for Sign Language Tutoring**" *MultiMedia, IEEE , Vol. 16, No. 1, pp. 81-93, January-March 2009.*
2. K. Imagawa, S. Lu, S. Igi, "**Color-based hands tracking system for sign language recognition**," *Automatic Face and Gesture Recognition. Proceedings. Third IEEE International Conference, Vol., No., pp. 462-467, April 1998.*
3. M. S. M. Asaari and S. A. Suandi, ‘‘**Hand gesture tracking system using adaptive Kalman filter**’’,*in Proc. 10th Int. Conf. Intell. Syst. Design Appl., Nov./Dec., pp. 166–171, 2010.*
4. E.-J. Holden, G. Lee, R. Owens, "**Automatic Recognition of Colloquial Australian Sign Language**," *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops, Vol. 2, No. , pp. 183188, January 2005.*
5. G. Awad, J. Han, and A. Sutherland, “**A Unified System for Segmentation and Tracking of Face and Hands in Sign Language Recognition**”, *18th International Conference on Pattern Recognition, 2006 (ICPR), Vol. 1, pp. 239-242, IEEE, 2006.*
6. S. Hadfield and R. Bowden, “**Generalized pose estimation using depth**”, *in Trends and Topics in Computer Vision, London, Springer, pp. 312-325, 2012.*
7. D. Uebersax, J. Gall, M. Van den Bergh and L. Van Gool, "**Real-time sign language letter and word recognition from depth data**" *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, Vol., No., pp. 383-390, November 2011.*
8. P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard and V. Athitsos, “**Comparing gesture recognition accuracy using color and depth information**”, *Proceedings of*

the 4th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 20:1-20:7, New York, ACM, 2011.

9. Mahmoud M. El-Gayyar, Amira S. Ibrahim, M.E. Wahed " **Translation from Arabic speech to Arabic Sign Language based on cloud computing**", *Egyptian Informatics Journal, 17, 295-303, 2016.*
10. Nada B. Ibrahim ,Mazen M. Selim, Hala H. Zayed**" An Automatic Arabic Sign Language Recognition System (ArSLRS)**", *Journal of King Saud University – Computer and Information Sciences 30, 470–477, (2018).*
11. AmitkumarShinde , Ramesh Kagalkar " **Sign Language to Text and Vice Versa Recognition using Computer Vision in Marathi** ", *International Journal of Computer Applications (0975 – 8887)*
12. Amruta L. Kabade, Dr.V.G. Sangam " **Canny edge detection algorithm",** *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 5, Issue 5, May 2016.*
13. Marek Hrůz · Pavel Campr · ErinçDikici · Ahmet Alp Kındıroglu " **Automatic fingersign-to-speech translation system**", *J Multimodal User Interfaces 4:61–79, 2011*
14. Allen J, Xu R, Jin J, "**Object tracking using camshift algorithm and multiple quantized feature spaces**". *In: Proceedings of the Pan-Sydney area workshop on visual information processing, vol 36. Australian Computer Society, Inc, pp 3–7, 2004.*
15. B. Kang, S. Tripathi and T. Q. Nguyen, "**Real-time Sign Language Fingempelling Recognition using Convolutional Neural Networks from Depth map**," *in arXiv: 50903001. 2015.*
16. L. Pigou. S. Dieleman, P.-J. Kindermans and B. Sehrauwen, "**Sign Language Recognition Using Convolutional Neural Networks**,"*in Computer Vision - ECCV 2014 Workshops, 2014.*
17. Garcia, Brandon and Viesca, Sigberto., "**Real-time American Sign Language Recognition with Convolutional Neural Networks**",*In Convoiurionai Neural Networks for Visnai Recognition at Stanford University, 2016.*
18. Moumen El-Melegy, Hashim Mokhtar. "**Incorporating prior information in the fuzzy C-mean algorithm with application to brain tissues segmentation in MRI**", *16th IEEE International Conference on Image Processing (ICIP), 2009.*
19. http://www.menasy.com/arab%20Dictionary%20for%20the%20deaf%202.pdf
20. Fang G, Gao W, Zhao D, "**Large vocabulary sign language recognition based on fuzzy decision trees**. IEEE Trans Syst Man Cybern Part ASyst Hum 34(3):305–314, 2004
21. Fels SS, Hinton GE (1993) Glove-talk, "**A neural network interface between a data-glove and a speech synthesizer**",*IEEE Trans Neural Netw 4(1):2–8*
22. Kong W, Ranganath S.,"**Signing exact english",** *modeling and recognition. Pattern Recognit 41(5):1638–1652,* 2008
23. Murakami K, Taguchi H., "**Gesture recognition using recurrent neural networks**",*In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 237–242,* 1991
24. Chen F, Fu CM, Huang CL, "**Hand gesture recognition using a real-time tracking method and hidden Markov models**",*Image Vis Comput 21(8):745–758,* 2003
25. Zaki MM, ShaheenSI.,"**Sign language recognition using a combination of new vision based features**",*Pattern Recognit Lett 32(4):572–577,* 2011
26. Assan M, Grobel K. "**Video-based sign language recognition using hidden markov models**", *In: International gesture workshop. Springer, pp 97–109*
27. Starner T, Pentland A., "**Real-time american sign language recognition from video using hidden markov models",** *In: Motion-based recognition. Springer, pp 227–243, 1997*
28. Wang RY, Popovic J.," **Real-time hand-tracking with a color glove**",*In: ACM transactions on graphics (TOG), vol 28. ACM, p 63,* 2009
29. Zhang Z, Huang F., "**Hand tracking algorithm based on superpixels feature**", *In: 2013 international conference on information science and cloud computing companion (ISCC-C). IEEE, pp 629–634,* 2013
30. Zou X, Wang H, Zhang Q., "**Hand gesture target model updating and result forecasting algorithm based on mean shift**", *J Multimed 8(1):1–8,* 2013
31. Jeyakar J, Babu RV, Ramakrishnan K., "**Robust object tracking with background-weighted local kernels**". *Comput Vis Image Underst 112(3):296–309,* 2008
32. Nadgeri SM, Sawarkar S, Gawande AD., **" Hand gesture recognition using camshift algorithm**",*In: 2010 3rd international conference on emerging trends in engineering and technology (ICETET). IEEE, pp 37–41,* 2010
33. Kian M., Alan W., Chin P., Shing C., " **Isolated sign language recognition using Convolutional Neural Network hand modelling and Hand Energy Image**", *Multimedia Tools and Applications, 78:19917–19944., 2019*
34. Chen S., "**Kalman filter for robot vision: a survey",***IEEE Trans Ind Electron 59(11):4409–4420,* 2012
35. Prince SJ. "**Computer vision: models, learning, and inference**", *Cambridge University Press,Cambridge,* 2012
36. Ross DA, Lim J, Lin RS, Yang MH., "**Incremental learning for robust visual tracking**", *Int J Comput Vis 77(1–3):125–141,* 2008
37. Wang D, Lu H, Yang MH., " **Online object tracking with sparse prototypes**", *IEEE Trans Image Process 22(1):314–325,* 2013
38. Zhong W, Lu H, Yang MH., " **Robust object tracking via sparsity-based collaborative model**", *In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1838–1845,* 2012
39. Zhou SK, Chellappa R, Moghaddam B., "**Visual tracking and recognition using appearance-adaptive**

models in particle filters", *IEEE Trans Image Process 13(11):1491–1506*, 2004

40. Shan C, Tan T, Wei Y., " **Real-time hand tracking using a mean shift embedded particle filter**", *Pattern Recognit 40(7):1958–1970*, 2007
41. Morshidi M, Tjahjadi T., "**Gravity optimised particle filter for hand tracking**", *Pattern Recognit 47(1):194–207*, 2014
42. Lionel P., Sander D., Pieter K., and Benjamin S., "**Sign Language Recognition Using Convolutional Neural Networks**", *ECCV 2014 Workshops, Part I, LNCS 8925, pp. 572–578, 2015*.
43. H. Almohamedh , S. Almotairi, "**Facial Emotion Recognition Using Eigenface and Feature Optimization"**, *International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No. 4, July-August 2019.*
44. J.Sutopo, M.Khanapi, M.A.Burhanuddin, Zulhawati1, "**Gesture Recognition of Dance using Chain Code and Hidden Markov Model**", *International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No. 6,Noveber-December 2019.*
45. M. Sardeshmukh1, V.Sardeshmukh, "**Performance Analysis of Human Detection and Tracking in Changing Illumination**", *International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No. 6,Noveber-December 2019.*
46. J. Robert, B. Del Rosario, "**Development of a Face Recognition System Using Deep Convolutional Neural Network in a Multi-view Vision Environment**", *International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No3, May-June 2019.*