Progression of Arabic Sign Language Recognition and Its Potential Advances

Noha A. Sarhan
College of Engineering and Technology
Arab Academy for Science and Technology
Alexandria 1029, Egypt

Yasser El-Sonbaty
College of Computing and Information Technology
Arab Academy for Science and Technology
Alexandria 1029, Egypt
Email: Yasser [AT] aast.edu

Sherine M. Youssef
College of Engineering and Technology
Arab Academy for Science and Technology
Alexandria 1029, Egypt

Abstract—**Being a primary means of communication amongst deaf people, Sign Language Recognition (SLR) has been drawing worldwide attention for decades. Only recently, research in Arabic Sign Language (ArSL) has witnessed a surge. The foremost goal of this paper is to demonstrate the progress in Arabic Sign Language Recognition (ArSLR), categorized according to acquisition method, and the foreseeable advances towards a more natural Human-Computer Interaction (HCI). The employed methods in current research are also critically analyzed with the intention of guiding future research towards developing a robust, convenient and applicable ArSLR system.**

Keywords: Arabic SLR; sensor-based SLR; vision-based SLR; HCI; Kinect; Image Processing.

I. INTRODUCTION

Since birth, humans naturally rely on the use of hand gestures in delivering a message and expressing themselves. Even as adults, they tend to unconsciously use hand gestures and body language to support what they are saying. Being a visually transmitted means of communication, signing has been the essential form of communication amongst people who are deaf or hard of hearing. Over the centuries, those gestures evolved to form complete, well-structured sign languages, which are made up of various hand gestures, arm movements, lip patterns and facial expressions [1]. The varieties of cultures lead to the development of different sign languages rather than having a universal Sign Language (SL).

Practicing SL amongst deaf people does not present a problem. However, the problem arises in imminent daily situations when the deaf are forced to communicate with the rest of the society who find difficulty in comprehending SL. In attempt to avoid frustration, they gravitate towards hiring an interpreter, or using writing tools for communication [2]. Nevertheless, both of the aforementioned solutions are far from ideal. Interpreters are not always immediately available, not to mention they are expensive and deny the involved parties their privacy. In the same manner, writing becomes aggravating in walking and standing situations. As a result, the deaf community became habituated to living in isolation, which in turn deprives them of many opportunities like receiving proper education, exploring their talents, career growth, and so forth.

Driven by the motivation to break barriers between all citizens, researchers in the field of (HCI) are recently focusing on developing automatic SLR systems [3]. The objective of such systems is to translate sign language to written text or speech. Inspired by the way humans interact with one another, researchers aim to make those systems as natural as possible. To facilitate two-way communication, a counterpart to this system would be one that converts text or speech into signs.

There are various ways SLR systems are classified. One approach is according to the type of gesture: static and dynamic. Static gestures can be represented by a single image, as the hand remains stationary. A dynamic gesture involves movement of the hand, thus represented in a video sequence where the hand shape or position changes throughout the duration of the sign. Another way to classify SLR systems is based on the type of the input: alphabets, isolated words or continuous sentences. Alphabets are usually represented by static gestures and seldom involve minor movement. Words are often presented by dynamic gestures, the input can either be isolated words or sentences performed by continuous signing. A different classification approach, the one considered in this paper, is based on the data acquisition device used: sensor-based and vision-based (or image-based). The first entails the user to wear gloves equipped with sensors that measure different parameters related to the gesture. On the other hand, vision-based systems rely on the use of a video camera that captures the performed gesture, sometimes aided with the user wearing simple colored gloves or without any external aid.

Being an example of hand gesture recognition, SLR is receiving growing attention worldwide for its value in helping the deaf and hard of hearing become more independent. For instance, recognition of continuous American SL using real-time HMM [4]; isolated words of Korean SL using fuzzy min-max neural network [5]; German SLR [6], Australian SLR [7] among others. Comparatively, research in ArSL is still in its infancy. A survey on current research trends in SLR that is presented in [8], where the general problems that should be tackled in order to increase the system's usability and applicability are explained. These issues are:

- Segmentation technique
- Unrestricted environment
- Size of dictionary
- Invariance
- Variety of gestures
- Generality
- Motion gestures
- Feature selection

The remainder of this paper is structured as follows; Section II highlights the characteristics of ArSL. Section III presents sensor-based ArSLR systems, followed by vision-based ArSLR systems in Section IV. Kinect is introduced in Section V along with SLR systems that use it, followed by a short description of the Leap Motion Controller and its application in ArSLR in Section VI. Section VII briefly mentions work done on developing complementary systems to ArSLR. Publicly available ArSL databases are demonstrated in Section VIII. Finally, Section IX concludes the paper.

II. Arabic Sign Language

The Arabic language is a widespread language and is the native language of countries of the Arab league and other neighboring countries, albeit in different dialects. In a similar manner, ArSL varies for every country, however efforts are being made towards unifying ArSL [9]. In spite of the dissimilarities, the gestures representing the Arabic alphabets (shown in Fig. 1) are the same [10, 11]. However, recognition of only alphabets is impractical as signers rarely spell out words since a gesture for every word already exists, with the exception of spelling out a name or an address. A survey on image-based ArSLR systems is present in [12], and a more thorough survey in [13] that includes sensor-based ArSLR systems as well.

The elements of the ArSL, almost like any other sign language, are: the hands, the mouth, the eyes, the face and the body. Different facial expressions, body postures, and lip patterns convey the meaning of the sign, the structure of the sentence, and the functionality of the word. Research focuses mainly on the hands, being the prime element to express the

Fig. 1. Gestures of Arabic sign language alphabets

word. In terms of the hands, the phonological structure lies in four elements:

- Hand configuration
- Hand orientation
- Articulation point
- Type of movement (if present, its direction, and its severity)

III. Sensor-based ArSLR

In this approach, the user is asked to wear special gloves that employ sensors, so as to acquire their hand movement. These sensors measure parameters such as hand's position, angle, finger flexion, and fingertips' locations. The collected data is then processed in order to recognize the performed gesture. Motion tracking devices are set up to collect the sensors' data, for example Flock of Bird (FOB), which is used in [14].

A thorough survey on different glove systems, their characteristics and applications can be found in [15]. Among the most suitable gloves for SLR are: Digital Entry Data Glove [16], Data Glove [17], [18], PowerGlove [17, 19], and CyberGlove [20]. The CyberGlove and PowerGlove (shown in Fig. 2) are the commonly used gloves in ArSLR [20, 21].

Fig. 2. Cyber Glove (left); PowerGlove (right)

Mohandes et al. [22] used the cost-effective PowerGlove as an interface between the system and the signer. The acquired measurements were the location (x, y, z), orientation (roll only) of the hands and finger flexure information of four fingers (no measurements for little finger). Three adults performed different signs. Due to the variation of the lengths of different signs, time division averages were computed by dividing the entire time of the sign into equal segments. Those averages are used as input to support vector machine (SVM) with two operation modes: training mode and recognition mode. Whilst the achieved recognition accuracy for 10 different signs was over 90%, increasing the dataset to 120 words resulted in below 70% accuracy. This is believed to be due to the limited measurements, low accuracy and imprecision of the PowerGlove.

In [14], Mohandes used the CyberGlove, which provides high-accuracy joint-angle measurements. Mohandes used two CyberGloves for recognition of two-handed Arabic signs along with two hand-tracking devices (Flock of Bird) to measure location (x, y, z) and orientation (yaw, pitch, roll) of each hand. A total of 56 measurements is provided for both hands from both devices (22 sensor signals from each Cyber Glove and 6 signals from each FOB). Time division into 10 segments was employed since different signs have different lengths. The mean and standard deviation were calculated giving a total of 20 values for each segment, leading to 1120 values to represent the signals from the 56 sensors. Accordingly, Principal Component Analysis (PCA) was applied for dimensionality reduction. For the recognition of signs, SVM was used with Radial Basis Kernel. Only one signer performed 100 two-handed signs to generate samples for the learning machine. When tested, recognition rate of 99.6% was achieved, however, the recognition rate did not exceed 63% when tested with another signer. Nonetheless, adding samples for the second signer in the training lead to 93% recognition rate. To obtain a signer-independent system, several signers should provide samples to the recognition system.

Mohandes and Deriche [23] developed a new approach aiming to enhance recognition performance of two-handed ArSLR. The authors used Dempster-Shafer (DS) Theory of evidence to investigate decision-level fusion, rather than data-level fusion. CyberGloves and FOB were also used, however they opted for Linear Discriminant Analysis (LDA) for dimensionality reduction. Minimum Distance (MD) classifier was used in two experiments, the first one used both CyberGloves only, while the second used FOB only. Finally, DS theory of evidence was used to combine the decisions from both experiments. An accuracy of 98.1% was achieved with fusion at decision level.

A summary of the aforementioned methods is illustrated in Table I, in terms of glove used, the acquired measurements, the classifier of choice, the dataset used for testing, and the corresponding recognition rate.

The advantage of using glove-based systems lies in their high levels of accuracy. Additionally, they do not require a special environment to work in nor are they affected by illumination changes. The most valuable advantage is bypassing the computationally expensive hand segmentation stage. Nevertheless, the need for cumbersome wired gloves imposes a great deal of inconvenience to the signer by confining their movement. In addition, the size of the glove might not be suitable for all hand sizes, which may in turn give inaccurate measurements. As a result, the system becomes significantly less natural than the way HCI is expected to be. Consequently, research shifted towards the more natural vision-based approach.

IV. Vision-Based ArSLR

Aiming to eliminate the restrictions caused by glove-based systems, researchers resort to vision-based techniques. They rely on the use of a video camera to capture the hand movement, thereby increasing naturalness of HCI. These recognition systems are generally comprised of five stages: image acquisition, image preprocessing, segmentation, feature extraction, and finally classification (Fig. 3). Vision-based gesture recognition systems are further divided into two categories: the first relies on the use of colored gloves; the second works on bare hands, which is more natural, but in turn more complex.

A. Vision-based ARSLR with External Aid

In this case, the user is asked to wear lightweight colored gloves. The gloves might be of solid color or with distinguished colors for each fingertip, as the one shown in Fig. 4 [24]. In comparison, this is more convenient than sensor-based methods, thereupon opted for in SLR.

In 2005, Assaleh and Al-Rousan [24] collected a dataset of 40 gestures representing 30 alphabets in the Arabic language. Thirty deaf participants were asked to wear gloves shown in Fig. 4. Prior to segmentation, the acquired RGB images were transformed to hue-saturation-intensity (HSI) color space. After locating the centroid of each identified region, thirty geometric features are extracted. These features are the vectors from the centroid of each region to the center of all other regions, and the angle between each of those vectors and the horizontal axis. The collected set of features is shown in Fig. 5 [24]. The performance using polynomial classifiers was compared to using adaptive neuro-fuzzy inference systems (ANFIS). The former proved to outperform the latter with a recognition rate of 93.41%.

TABLE I. SUMMARY OF SENSOR-BASED ARSLR TECHNIQUES

Method	Glove Type	Acquired measurements	Classifier	Dictionary size	Recognition rate
Mohandes *et al.* [22] (2004)	PowerGlove	• Hand location (x, y, z) • Hand orientation (roll only) • 4 finger flexure information	SVM	10 words	90%
				120 words	20%
Mohandes [14] (2013)	Cyber Glove	• Hand location (x, y, z) • Hand orientation (yaw, pitch, roll)	SVM	100 words	99.6%
Mohandes & Deriche [23] (2013)	Cyber Glove	• Hand location (x, y, z) • Hand orientation (yaw, pitch, roll) • LDA for dimensionality reduction	MD Classifier	100 words	98.1%

A reduction of 57.4% in the number of misclassifications was achieved when using the polynomial classifiers rather than ANFIS. The proposed algorithm is robust to lighting conditions, by converting to HSI color space, and also invariant to distance between the signer and the camera. However, further improvements can be obtained when using the polynomial classifier by compensating for the prior probabilities, given the distribution is not uniform. Also, the algorithm is does not consider both hands

Working with the same gloves and extracting the same features as those in [24], Maraqa and Abu-Zaiter [25] aimed to bring in the use of recurrent neural networks in hand gesture recognition in 2008; the Elman recurrent network and a fully recurrent neural network. On a database of 900 samples representing 30 gestures, his fully recurrent neural network proved superior, with an accuracy rate of 95.11%, compared to 89.66% accuracy using Elman network. Later in 2012, the authors extended their work in [26] by testing different types of neural networks. The fully recurrent neural network still gave the highest recognition accuracy.

Fig. 3. Stages of vision-based SLR

Mohandes *et al.* [27] opted for yellow and orange gloves with the restriction of being different from the background color. First of all, the signer's face is detected using Gaussian skin model in chromatic color space. The centroid of the face is used as a reference point for the hand movement. Region-growing technique was used to track the hands in the RGB color image. Afterwards, prominent geometric features are extracted, which are: centroid of both hands with respect to the centroid of the face; eccentricity of bounded ellipse for both hands; the angle of the first principal component for both hands; and the area of both hands. Eccentricity is calculated using eigenvalues of the following matrix:

$$\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \quad (2)$$

where,

$$a = \iint_{l'} (x')^2 \, dx' dy' \quad (3)$$

$$b = \iint_{l'} x'y' \, dx' dy' \quad (4)$$

$$c = \iint_{l'} (y')^2 \, dx' dy' \quad (5)$$

HMM is used for classification; hence, time-varying sequences need not be of equal lengths. Using a dictionary of 50 signs, the achieved recognition accuracy was 98%. In a successful attempt to tackle one of the main challenges of developing SLR system, the same authors later extended their work in [28] on a larger dictionary of 300 signs achieving 95% recognition accuracy. Although this algorithm gives high recognition rates, even when working on large dictionaries, face detection and region growing algorithm are both computationally expensive.

Fig. 4 Colored gloves with different markings at fingertips and wrist

In [29] Shanableh and Assaleh, 2007, worked on achieving a user-independent ArSLR system. Three participants performed 23 different Arabic words/phrases 50 times over 3 different sessions. With the exception of wearing colored gloves, no other restrictions were imposed on signer's clothing or image background. Taking advantage of the color gloves, the video sequences were segmented in the RGB color space. Afterwards, temporal domain and spatial domain features were extracted. Weighted directional accumulated image differences represented the motion in the video sequences, thereby eliminating the temporal dimension:

$$AD_{g,j} = \sum_{j=1}^{n-1} \partial\left(\left|I_{g,j}^{(j)} - I_{g,i}^{(j-1)}\right|\right) \qquad (1)$$

Where n is the total number of images in the i^{th} repetition of a gesture at index g. ∂_j is a binary threshold function of the j^{th} frame. Spatial features were subsequently extracted by transforming the absolute difference images to the frequency domain using Discrete Cosine Transform (DCT) followed by zonal coding to form the feature vectors. Lastly, k-nearest neighbor (KNN) was employed with a classification rate of 87% in user-independent mode.

In the same year, Shanableh and Assaleh extended their previous work in [30] and [31]. The authors compared the performance of Bayesian and KNN with Hidden Markov Model (HMM)-based classification. The main advantage of this algorithm is that it omits the temporal dimension, and therefore there is no need to use a time-sensitive classifier.

Once again, the authors in [32] presented a solution for user-independent recognition of isolated ArSL gestures. The dataset consisted of 3450 samples covering 23 isolated gestures from 3 signers. Accumulated difference images are used to eliminate the temporal-domain, then extraction of spatial-domain features proceeds. DCT is employed to transform difference images to frequency domain, followed by Zonal coding. KNN and polynomial classifiers were employed to validate the suggested solution, reaching a classification rate of 87%

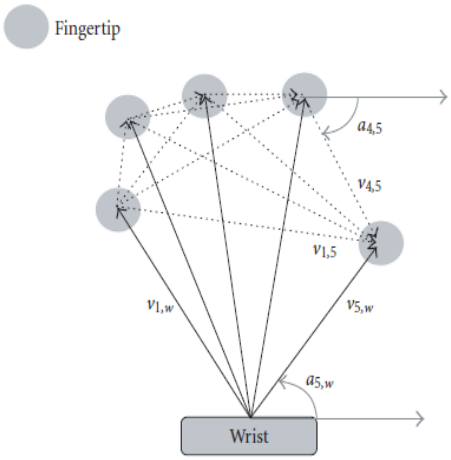

Fig. 5 Vectors representing the collected set of features

In [33] Naoum et al. presented a new recognition system for Arabic alphabets by extracting the vertical and horizontal profiles of the images. The acquired image is clipped and narrowed a priori so that it becomes scale-invariant, followed by an image masking process. Two histograms are generated for each character expressing the vertical and horizontal behaviors. In the end, the histogram is used to detect the surface behavior using KNN algorithm. Testing with different glove colors, the realized hit rates are 50% for the bare hands; 75% wearing a red glove, 65% wearing a black glove, and 80% wearing white gloves.

In the recent work by Mashagba et al. [34] the use of Time-Delay Neural Networks (TDNN) in an isolated-word ArSLR system was demonstrated for the first time. Yellow and blue gloves were used for the right and left hands respectively, with the constraint that the signer's clothing and background be of different colors. The collected video sequences are converted into frames, followed by color segmentation using Gaussian Mixture Model (GMM). The user selects the regions of interest, and then the mean and covariance of each colored segment are calculated. Next, the following geometric features were chosen to describe the two-dimensional projection of each hand: position of the centroid of each hand with reference to upper area of each frame, the horizontal and vertical velocities of both hands, and the area of each hand. TDNN was employed to classify the gestural patterns owing to its success in learning and classifying spatio-temporal patterns. Based on the experiments with 40 ArSL gestures, the total average recognition rate is 70.0%.

Table II summarizes the preceding methods, highlighting the type of features extracted, dataset size, classifier, and the achieved recognition rates.

This approach presents a step towards a more natural interaction. In addition, the colored gloves lessen the complexity of the segmentation process especially in the case of hands over face occlusion [35]. Ensuring that the glove color differs from signer's clothes and background, further simplifies segmentation. On the other hand, processing complexity is increased and the signer is not totally

TABLE II. SUMMARY OF VISION-BASED ARSLR TECHNIQUES WITH EXTERNAL AID

Method	Type of extracted features	Dictionary size	Classifier	Recognition rate
Assaleh & Al-Rousan [24] (2005)	Geometrical features	30 alphabets	Polynomial classifier	90%
			ANFIS	57.4%
Maraqa & Abu-Zaitar [25] (2008)			Elman recurrent network	95.11%
			Fully recurrent neural network	89.66%
Marqa & Abu-Zaitar [26] (2012)			Feed-forward neural network	79.33%
			Jordan neural network	84.56%
Shanableh & Assaleh [29] (2007)	Spatial and temporal features	23 words/phrases	KNN	87%
Shanableh & Assaleh [31] (2007)			Fisher's linear discriminants	About 95%
Mohandes *et al.* [27] (2005)	Geometrical Features	50 words	HMM	98%
Mohandes *et al.* [28] (2012)		300 words		95%
Naoum *et al.* [33] (2012)	Profile of the image	Alphabets	KNN	80%
Mashagba *et al.* [34] (2013)	Geometrical features	40 words	TDNN	70.0%

comfortable. Moreover, some restrictions remain enforced on the environment.

B. Vision-Based ArSLR Without External Aid

In order to provide the utmost comfort and naturalness, the signer is freed from using any gloves. Recognition is done using images of bare hands, where skin color is detected. Image processing techniques [67, 68] can also help in compressing and matching the output images.

In 2001, Al-Jarrah and Halawani [36] were able to automatically translate 30 manual gestures of the Arabic alphabets without the assistance of any gloves. The image segmented using the iterative thresholding algorithm. Employing a border-tracing algorithm, at which point only the contour of the hand is of interest, identifies the border of the gesture. Next, the features are extracted by calculating the lengths of vectors originating from the center of the hand to the fingertips. In the end, the gesture is recognized using Adaptive Neuro-Fuzzy Inference Systems (ANFIS) fulfilling 93.55% recognition accuracy. The strength of this algorithm lies in the extracted features, as they are scale, translation and rotation invariant. Therefore, the system is robust to changes in position of the signer, and size of the hands.

Prior to their work in [32], Shanableh *et al.* ss[37], 2007, had presented an isolated ArSLR system that didn't rely on the use of colored gloves in 2007. They extracted temporal features from the image via accumulation of forward, backward, or bi-directional prediction errors of the input video sign. The prediction error is then thresholded into a binary image representative of the temporal features, thereby eliminating the temporal dimension. Spatial feature extraction then follows where two approaches were examined. The first entails frequency domain transformation on the accumulated temporal differences, followed by Zonal coding. The second is projecting the pixel intensities of the accumulated difference through Radon transformation followed by an ideal low-pass filter. Simple classification techniques, namely KNN and Bayesian classification, were compared to HMM-based schemes yielding comparable results since the temporal dimension was eliminated and therefore alleviating the need for computationally expensive classifiers.

In 2010, Hemayed and Hassanien [38] presented an instinctive practical setup that converted the recognized ArSL alphabets into voice, but, alas, not in real-time. Using a skin profile, skin is detected in YCbCr color space, followed by morphological operations resulting in a binary image with the hand and head blobs evaluated as foreground (FG). Next, Prewitt operator is used to detect edges of the hand blob - the chosen feature to represent the hand shape. PCA is used for dimensionality reduction. KNN is used in the classification phase leading to 97% recognition accuracy. Even though comparing images to stored templates is robust against small differences in hand size, it is sensitive to changes in hand orientation. Moreover, because of the way

the hand is differentiated from the head, the signer must be right-handed. In addition, high and poor illumination conditions have a negative effect on recognition.

El-Bendary *et al.* [39] proposed a versatile Arabic Sign Language Alphabets Translator (ArSLAT) that is able to recognize words and sentences as well in 2010. The captured gesture undergoes five stages. First, video segmentation into frames. Second, best frame representing the letter is detected. Third, the sign is categorized according to the position of the wrist in the image. Fourth, the features are extracted using the centroid as reference, the distance to the edges of the hand are calculated as a 50-D vector. Finally, Minimum distance classifier (MDC) and multilayer perceptron (MLP) neural networks were used to measure the performance of the ArSLAT system achieving an accuracy of 91.3% and 83.7% respectively. The strength of ArSLAT lies in the choice of rotation, translation, and scale invariant features. Additionally, the category detection phase doesn't only raise recognition accuracy, but also reduces processing time by decreasing the matching operation. On the other hand, the signer has to spell out the word rather than use its already existent gesture, which might be burdensome for everyday life.

Assaleh *et al.* [40] adopted two-tier spatio-temporal feature extraction for recognition of continuous ArSL. First, accumulated differences (ADs) of consecutive frames represent motion. Second, Discrete Cosine Transform (DCT), followed by Zonal coding to construct the feature vector. HMM was the chosen classifier. Two different databases were involved. The first was for isolated gesture recognition, where three signers performed 23 gestures a total of fifty times over three different sessions. The second was for continuous recognition, where 40 sentences were created from an 80-word lexicon without enforcing any restrictions on grammar or sentence length. The isolated gestures were concatenated and tested, and as expected, gave better average recognition rates as compared to continuous gesturing. The former gave an average of 96% sentence recognition and 98% word recognition, while an average of 75% sentence recognition and 94% word recognition were reached from the latter.

Working on bare hands in 2011, Zaki and Shaheen [41] selected three features based on the four components of sign languages. The proposed system starts by detecting the hands using a skin color detector, followed by connected component labeling where three components portraying the two hands and the face are detected. The selected appearance-based features are:

- Principal Component Analysis (PCA): configuration & orientation of the hand
- Kurtosis position: articulation point

$$Kurt(x) = \frac{E(x-\mu)^4}{\sigma^4} \qquad (6)$$

  where, x is the image, μ is the mean of x, and σ is the standard deviation of x

- Motion chain code: hand motion

HMM is employed for classification for its ability to process data of variable temporal lengths. With a database of 50 signs of ASL, the recognition error rate was 10.90%. The weakness of this algorithm is that it relies on a skin color detector, which imposes on constraints on the background, clothing, and is affected by other visible body parts such as the face, and the neck. However, the use of motion chain code, is easy and robust.

Youssif *et al* [42], 2011, granted the signer an unconstrained environment; the signers were glove-free, with no restrictions on clothing, background, skin color, age, and gender. The system they brought in recognized 20 isolated ArSL words. Circles and ellipses were used to represent the hand blob and ridge features. Skin detection is done in HSV color space, followed by Canny edge detection. Among the selected features are: head position, coordinates of the center of the hand, direction angle of hand, image motion extracted from the variations of image intensities over time, and corner points with big eigenvalues suitable for tracking. HMM was utilized, achieving 82.2% accuracy using only eight features. Working with HSV color space makes the algorithm invariant to illumination changes.

Albelwi and Alginahi [43], 2012, present a real-time vision-based system that recognizes static ArSL alphabets. The hand is tracked in successive frames using Haar-like classifier. Next, skin is detected in HSV color space, thereby eliminating the effect of illumination changes. The selected features are shift and rotation invariant, and reflect the phonetic structure of the sign language. Therefore, contour of the hands is extracted which only reveals information about the shape boundary, ignoring interior information. Fourier transform is then applied on the shape signature generating the Fourier descriptors of the shape. Eventually, KNN was used for classification, and 90.55% accuracy rate was achieved. A strong point about the proposed system is that processed test images that got a recognition rate exceeding 90% are assigned as new training images. This provides a means of covering all variants of the hand and enhances the scalability of the recognition system.

In 2013, Elons *et al.* [44] offered a novel approach to handle pose variations in 3-D object recognition. Two cameras were positions at two different viewing angles 90° apart. The cameras are rotated 90° and an image is acquired from each camera at every 5° increment. A total of 19 3-D image sets are acquired; 10 of which are used for training and the remaining 9 sets for testing. Features are then generated from the captured images using Pulse Coupled Neural Network (PCNN) Feature Generation module. Feature selection was then carried out. Working in an unconstrained environment with uneven lighting and background and a dataset of 50 isolated words, a recognition accuracy of 96% was achieved.

Table III sums up the aforementioned methods used in vision-based ArSLR without any external aid. The type of features, dataset size, classification method, and corresponding recognition rates are highlighted.

Despite reaching ultimate convenience, extracting the hand from the image would require vast computations. The

use of a skin color model would detect other areas such as face, neck and possibly arms, making it difficult to extract the hand. Additionally, the wide range of skin colors makes it even harder to reach a signer-independent recognition system. With all vision-based techniques (with or without external aid), illumination changes and occlusions present the most challenging problem, as a result increasing the system complexity and likely lowering recognition accuracy.

V. KINECT IN SLR

In collaboration with Microsoft, Prime Sense developed the Kinect sensor. It embodies an RGB camera, a depth sensor, an accelerometer, a motor, and a multi-array microphone. The depth image is represented by a 2D matrix with depth of each pixel in the scene. The RGB and IR cameras both capture data at 640x480 pixels at a rate of 30fps, thereby acquiring synchronized color and depth images, namely RGB-D (RGB + depth) images. The depth data make it possible to obtain the skeleton of humans in front of the sensor and accurately locate 20 joints [45]. The embedded software enables tracking of up to two skeletons, invariant to pose, body shape, clothing, etc. Due to its low cost and availability, the Microsoft Kinect has been used in many research areas including: object tracking and recognition, human activity analysis, and hand gesture analysis [46].

Kinect has been exploited in hand gesture analysis owing to its many advantages. It tackles the most critical problem faced in employing a vision-based approach, that is hand detection. Locating the hand using a skin color technique or markers color is not robust against changes in lighting conditions, however, using the depth information provided by Kinect, segmentation becomes illumination-invariant and is no longer affected by cluttered backgrounds, clothing or other body parts (e.g. neck and face which would be detected if skin color based segmentation is used). Advantages of Kinect are not only limited to offering a non-restricted environment, but it also provides a natural human-computer interaction. Moreover, the use of skeletal data facilitates locating the hand robustly in the image using the hand joints coordinates.

Unfortunately there are some problems with the data acquired from Kinect, especially the depth data. Both the color and depth images have noises due to the low resolution of the cameras [46]. Additionally, the depth image suffers from holes where some pixels observed by the RGB camera have no corresponding depth information [45]. Some suitable filtering techniques were presented in [45]. Furthermore, partial or full occlusion of the hand remains a challenge.

Recently, Kinect has been employed in various SLR systems. Several tests have been made on different sign languages. These include, but are not limited to: Polish SL (PSL) [47], Brazilian SL (BSL) [48], Chinese SL [49] [50], among others [51] [52]. Table IV summarizes the achieved recognition rates for the aforementioned systems, and their corresponding dataset size. The techniques used in each of the above systems are discussed below.

In 2013,Oszust and Wysocki [47] tested the use of two sets of features, the first represents the 3-D positions of the skeletal joints acquired from Kinect, and the second set describes the hand shape. In the latter, a combination of skin color model along with depth information about the objects were used to segment the hands. The extracted features for the hands are: center of gravity, area, compactness, eccentricity, depth difference between hand and face, and orientation. Dynamic Time Warping (DTW) was applied to align time series of different lengths. Using 30 PSL words, results using ten-fold cross-validation test yielded 89.33% accuracy with the first feature set, and a higher accuracy of 98.33% using the second feature set.

TABLE III. SUMMARY OF VISION-BASED ARSLR TECHNIQUES WITHOUT EXTERNAL AID

Method	Type of extracted features	Dictionary size	Classifier	Recognition rate
Al-Jarrah & Halawani [36] (2001)	Geometrical features	30 alphabets	ANFIS	93.55%
Zaki & Shaheen [41] (2011)	Representative of phonetic structure of SL	50 words	HMM	98.1%
Youssef *et al.* [42] (2011)	Geometrical Features	20 words	HMM	82.2%
Hemayed & Hassanien [38] (2010)	Structural and geometrical features	Alphabets	KNN	97%
El-Bendary *et al.* [39] (2010)	Geometrical features	Alphabets translator	MDC	91.3%
			MLP	83.7%
Assaleh *et al.* [40] (2010)	Spatial and temporal features	23 concatenated words	HMM	96%
		40 sentences; from 80-word lexicon		75%
Albelwi & Alginahi [43] (2012)	Representative of phonetic structure of SL	Alphabets	KNN	90.55%

For scalability purposes, in 2014 Almeida *et al.* [48] extracted features representing the phonological structure of BSL. Seven features were obtained relating to the structural elements of BSL. The videos were summarized a priori, by applying a clustering technique, in order to reduce the number of redundant frames. The extracted features are: 2-D distance between hands and shoulder center, 3-D distance between signer and camera, velocity of each sign using optical flow technique, area of the hands, corners' average position using Harris corner detector, detected lines using Hough transform, and amount of common points between frames using the descriptor algorithm SURF (Speed-Up Robust Features). Finally SVM was used for classification, achieving an average accuracy of 80% with a dataset of 34 words.

In 2013, Agarwal and Thakur [50] used a simple technique to recognize the gestures of the ten digits in Chinese SL. The entire processing was on the stream of depth images only. Images were first denoised using Gaussian filter, eroded, and then the background was subtracted. For every gesture, the depth and motion profiles were captured; a technique used in [53]. Using multi-class SVM classifier, 92.31% accuracy was achieved.

Focusing on the right hand, Geng *et al.* [49], 2014, were able to recognize 20 gestures from the Chinese SL by acquiring 3D skeleton joints data from Kinect, and capturing the hand location and its 3D trajectories in spherical coordinates. Due to their relative kinematic connectivity, 3D trajectories of the wrist and elbow were also obtained. Using Extreme Learning Machine for testing, a recognition rate of 69.32% was attained.

TABLE IV. SUMMARY OF SLR SYSTEMS USING KINECT

Method	Dictionary size and language	Recognition rate
Oszust & Wysocki [47] (2013)	30 Polish words	98.33%
Moreira *et al.* [48] (2014)	34 Brazilian words	Individually above 80%
Geng *et al.* [49] (2014)	20 Chinese words	69.32%
Agarwal & Thakur [50] (2013)	10 Chinese digits	92.31%
Martinez [54] (2011)	14 homemade words	95.238%

In 2011, Capilla [54] developed a generic SL translator that invariant to the SL used. The user is able to train the system and add new words to the dictionary. The focus was on the joints of both hands and elbows, expressed once in Cartesian coordinates as well as in the spherical coordinates. The data is normalized in order to become invariant to the user's position and size, providing a more robust system. Nearest-Group classifier with DTW and Nearest-Neighbor with DTW were used for testing. For a dictionary of 14 homemade signs, the system achieves an accuracy of 95.238%.

A. ArSLR using Kinect

In 2015, Sarhan [69] proposed a system that combined skeletal data and depth information for hand tracking and segmentation obtained from Kinect, without relying on any color markers, or skin color detection algorithms. The extracted features describe the four elements of the hand that are used to described the phonological structure of ArSL: articulation point, hand orientation, hand shape, and hand movement. Hidden Markov Model (HMM) was used for classification using ten-fold cross-validation, achieving an accuracy of 80.47%. Singer-independent experiments resulted in an average recognition accuracy of 64.61 %.

Fraiwan [70] used Kinect to capture the Arabic sign language gestures and transformed them into Arabic text, which in turn can be translated into any spoken language. Web services were used to generate the spoken sounds. They used hand and fingers identification and motion recognition in their algorithm. The accuracy in identifying the implemented characters was shown to exceed 80%.

VI. LEAP MOTION CONTROLLER IN SLR

The Leap Motion Controller (LMC) is a sleek motion-tracking device that gives a robust hand model [55]. In a very recent study, Mohandes *et al.* [56] introduced the use of LMC for recognition of ArSL. The twenty-eight static, single-handed alphabets of the Arabic alphabets are recognized. The LMC does not deliver images; the driver software processes the acquired data and returns 23 features for every frame. The authors opted for 12 features pertinent to ArSL. These include: finger length, finger width, average tip position with respect to x-, y-, and z-axis, hand sphere radius, palm position with respect to x-, y-, and z-axis, hand pitch, roll and yaw. For each letters 10 samples were acquired, each composed of 10 frames from which the mean of each feature is calculated in order to analyze relevance of extracted features. The performance of two classifiers was compared; naive Bayes classifier and multilayer perceptron neural network. An overall accuracy of 98.3% with the former and 99.1% with the latter were achieved. Despite the promising results, some of the signs were still misclassified. It is believed that is due to occlusion of fingers by others resulting from using one LMC from one side. The authors intend to explore the effect of using two LMC units placed at different positions.

VII. COMPLEMENTARY SYSTEMS

Research in ArSL is not only limited to translating signs into text or speech, several endeavors have been done in order to develop complementary systems that convert text or speech into signs using 3D hand models or synthesized avatars [57-61]. Furthermore, in attempt to achieve greater accuracy, error detection and correction techniques using a semantic-oriented approach, where semantic-level errors and

lexical errors are corrected, as a post-processing step were developed [62]. For portability, ArSLR mobile phone applications have been developed [63-65, 10].

VIII. Arabic Sign Language Datasets

Despite Arabic being a widespread language, there are very few organizations for ArSL. We aim to present a list of existing ArSL datasets in order to encourage future research.

The first labeled and segmented ArSL dataset was collected by the authors in [40] can be made available upon request. There are two different databases:

1. Isolated gestures
2. Continuous sentences

Three signers perform 23 gestures chosen from a greeting session that make up the first dataset. Each signer performed each gestures 50 times, a total of 150 gestures

The second dataset is composed of 40 sentences made up from an 80-word lexicon. Each sentence was performed 19 times by just one singer.

Another ArSL database [42] was collected using a single video camera in AVI format. The database included only 20 gestures of isolated ArSL words, each repeated 45 times. Different signers performed the gestures, and no restrictions were imposed on the singer's clothing nor the background. The signers used their bare hands without wearing any colored gloves or markers.

The aforementioned datasets is that they cover a very few words. Another problem is that the image quality is poor, in addition, neither of the databases include non-manual signs that involve the lips, facial expressions, and body language.

Only recently, SignWorld Atlas [66], a benchmark dataset for ArSL has been made publicly available. The database contains images and videos of both manual, and non-manual signs. Ten singers contributed in making the DB that contains about 500 elements. The database contains:

- Hand shapes in isolation and in single signs
- The Arabic finger spelling alphabets
- Numbers
- Movement in single signs
- Movement in continuous sentences
- Lip movement in Arabic sentences
- Facial expressions

Fig. 6 shows a sample image from each of the above elements in the database. The authors explain in details how the database is organized. The authors also tested the dataset for its efficiency.

It is noteworthy, that the publicly available datasets are not applicable to all technologies. For instance, none of these dataset is applicable for use with Kinect, since they were all captured using regular video cameras that do not acquire any depth information.

IX. Conclusion

With the ongoing advancements in the fields of HCI and gesture recognition, creating a natural and convenient interface for the deaf to integrate with other citizens became an attainable goal. Relying on computer vision also makes it affordable and user-friendly. Nevertheless, with the persisting open problems, there still remains much room for improvement.

The main challenge in ArSL is the lack of existence of well-organized documentation for ArSL in many countries. ArSL differs across different countries that speak the language, due to their different cultures. Not much attention has been paid towards unifying ArSL so far. Collecting a dataset can be a challenge as it requires personal contacts with such organizations.

As a result, all current research in ArSL work on small subset of the dictionary, which questions the practicality and

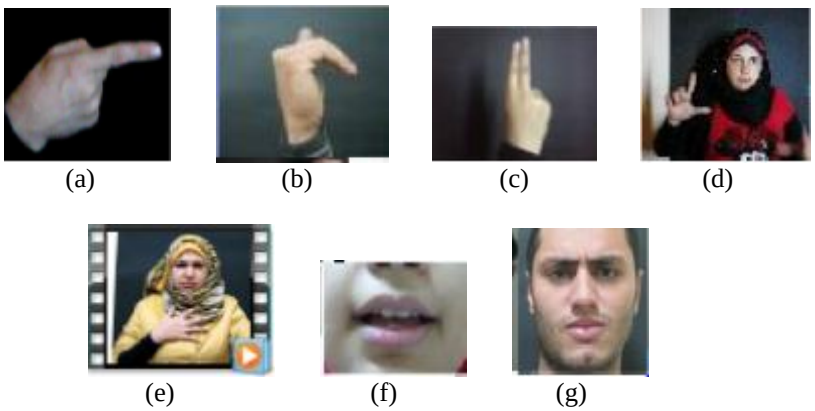

Fig. 6 SignsWorld Atlas Snapshots of ArSL: (a) Hand shapes; (b) Arabic alphabets; (c) Numbers; (d) Individual signs; (e) Continious sentences videos; (f) Lip patterns; (g) Facial expressions

applicability of such systems on a larger scale. Additionally, the ability of these systems to work in real-time is fundamental, especially in emergencies such as in hospitals or in cases of accidents, also to prevent frustration in everyday situations.

As demonstrated in this review, several approaches to ArSLR exist, each having its advantages and disadvantages.

The sensor-based approach significantly reduces any restrictions on the environment. In addition, they provide high level of accuracy, and bypass the most critical steps in other approaches: hand detection and segmentation. On the other hand, choosing the most appropriate glove is not an easy task; the number of sensors on the gloves need to be considered, size of the gloves and calibration and it could not be suitable for different hand sizes, which may in turn give inaccurate results Furthermore, the use of such cumbersome gloves is very inconvenient to the signer.

In the vision-based approach is more attractive in terms of naturalness and HCI. However many issues still remain unsolved. The use of colored gloves still burdens the user with having to equip oneself before using the system, however it makes the segmentation process easier, as locating the hands becomes less of a challenge. In addition, the use of colored gloves always imposes restrictions on signer's clothing and the background. Also, extreme lighting conditions might affect segmentation.

Although skin color based approaches alleviate many of these restrictions, it still has many inherent problems. Different visible body parts such as the neck, face, and arms might affect segmentation process. Illumination variations, background colors and the wide range of skin colors also make hand extraction a difficult task.

In all vision-based approaches hand occlusion remains one of the biggest challenges. One must ensure that both hands are visible at all times. If one hand overlaps the other, or the face, segmentation would be very challenging. Also, hand tracking throughout the gesture should be robust to changing the position of both hands; some gestures involve crossing both hands over one another.

The availability of inexpensive devices such as the Microsoft Kinect and Leap Motion Controller, ultimate naturalness and faster processing are now reachable, thereby increasing feasibility of a robust, practical and convenient ArSLR system. However, to the best of our knowledge, no dataset in ArSL using such devices currently exists.

In conclusion, research in ArSL is in its infancy and has not been employed on a large scale yet (only covers a vocabulary of less than 300 signs). Furthermore, for the system to be practical, research should focus on recognition of continuous ArSL. These systems would be representative of real-life situations. The main challenge is how to detect the beginning and end of every gesture. The few work on continuous ArSLR can be found in [27, 29] and [40,43,44]. The foremost goal still remains, that is developing robust system that does not impose restrictions on the signer nor the environment, and gives accurate result.

REFERENCES

[1] W. Jiangqin et al.: A simple sign language recognition system based on data glove. Fourth International Conference on Signal Processing Proceedings (1998) . doi:10.1109/ICOSP.1998.770847

[2] M. P. Paulraj, S. Yaacob, M. S..Z. Azalan, and R. Palaniappan, A phoneme based sign language recognition system using 2D moment invariant interleaving feature and Neural Network, IEEE Student Conference on Research and Development (SCOReD) (2011), doi:10.1109/SCOReD.2011.6148718.

[3] M. E. Al-Ahdal, and N. M. Tahir, Review in Sign Language Recognition Systems. IEEE Symposium on Computers & Informatics (ISCI) (2012). doi:10.1109/ISCI.2012.6222666.

[4] T. Starner, J. Weaver, and A. Pentland, Real-time American sign language recognition using desk and wearable computer based video. IEEE Transactions on Pattern and Machine Intelligence (1998). doi: 10.1109/34.735811 .

[5] J. S. Kim, J.-S, W. Jang, and Z. Bien, A dynamic gesture recognition system for the Korean sign language (KSL). IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics (1996). doi:10.1109/3477.485888.

[6] B. Bauer, and H. Hienz, Relevant features for video-based continuous sign language recognition. IEEE International Conference on Automatic Face and Gesture Recognition (2000). doi:10.1109/AFGR.2000.840672 .

[7] E. J. Holden, G. Lee, and R. Owens, Automatic Recognition of Colloquial Australian Sign Language. Application of ComputerVision, 2005. Seventh IEEE Workshops on WACV/MOTIONS'05 (2005). doi: 10.1109/ACVMOT.2005.30

[8] S. Kausar, and M. Y. Javed, A Survey on Sign Language Recognition. Frontiers of Information Technology (FIT) (2011). doi: 10.1109/FIT.2011.25

[9] M. Jemni, Toward the creation of an Arab Gloss for arabic Sign Language annotation. Fourth International Conference on Information and Communication Technology and Accessibility (ICTA) (2013). doi: 10.1109/ICTA.2013.6815292

[10] S. M. Halawani, Arabic Sign Language Translation System on Mobile Devices. International Journal of Computer Science and Network Security (IJCSNS). pp.251-256, 2008.

[11] " القاموس الاشاري العربي للصم". [Online]. Available at: http://www.menasy.com/arab%20Dictionary%20for%20the%20deaf%202.pdf

[12] M. Mohandes, J. Liu, and M. Deriche, A survey of image-based Arabic sign language recognition. 11th International Multi-Conference on Systems, Signals & Devices (SSD) (2014). doi: 10.1109/SSD.2014.6808906.

[13] M. Mohandes, M. Deriche, and J. Liu, Image-based and sensor-based Approaches to Arabic sign language recognition. IEEE Transactions on Human-Machine Systems (2014). doi: 10.1109/THMS.2014.2318280.

[14] M. A. Mohandes, Recognition of two-handed Arabic signs using the CyberGlove. Arabian Journal for Science and Engineering (2013), pp.669-677.

[15] L. Dipietro, A. M. Sabatini, and P. Dario, A Survey of Glove-Based Systems and Their Applications. IEEE Trans. Systems, Man and Cybernetics, Part C: Applications and Reviews (2008). doi: 10.1109/TSMCC.2008.923862.

[16] G. Grimes, Digital data entry glove interface device. U.S. Patent 4414 537, AT&T Bell Lab., Murray Hill, NJ, Nov. 1983.

[17] J. j. LaViola, A survey of hand posture and gesture recognition techniques and technology. Brown Univ., Providence, RI, Tech. Rep. CS-99-11, Jun. 1999.

[18] H. Eglowstein, Reach out and touch your data. *Byte*, vol. 15, no. 7, pp. 283-290, 1990.

[19] S. A. Mascaro, and H. H. Asada, Photoplethysmograph fingernail sensors for measuring finger forces without haptic obstruction. IEEE Transactions on Robotics and Automation (2001). doi:10.1109/70.964669

[20] (Oct. 2013). CyberGlove. [Online]. Available at:http://www.cyberglovesystems.com/products/cyberglove-II/photos-video

[21] PowerGlove. [Online]. Available at: http://www.pcauthority.com.au/News/343603,5-of-the-maddest-game-controllers.aspx.

[22] M. Mohandes et al., Automation of the Arabic sign language recognition. International Conference on Information and Communication Technologies: From Theory to Applications (2004). doi: 10.1109/ICTTA.2004.1307840.

[23] M. Mohandes, and M. Deriche, Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence. In Proceedings Computing, Communications and IT Applications Conference (ComComAp) (2013). doi: 10.1109/ComComAp.2013.6533615.

[24] K. Assaleh, Al-Rousan, M.: Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers. EURASIP Journal on Applied Signal Processing (JASP), pp. 2136-2145, 2005.

[25] M. Maraqa, and R. Abu-Zaiter, Recognition of Arabic Sign Language (ArSL) using recurrent neural networks. Applications of Digital Information and Web Technologies (2008). doi: 10.1109/ICADIWT.2008.4664396.

[26] M. Maraqa, Recognition of Arabic sign language (ArSL) using recurrent neural networks. J. Intell. Learn. Syst. Appl., pp. 41-52, 2012.

[27] M. Mohandes, and M. Deriche, Image based Arabic sign language recognition. Proceedings of the 8th International Symosium on Signal Processing and its Applications (2005) pp. 86-89, Aug. 2005.

[28] M. Mohandes, A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a hidden Markov model. Computers and Electrical Engineering (2012). doi: 10.1016/j.compeleceng.2011.10.013.

[29] T. Shanableh, and K. Assaleh, Arabic sign language recognition in user-independent mode. International Conference on Intelligent and Advanced Systems, 2007. ICIAS (2007). doi: 10.1109/ICIAS.2007.4658457 .

[30] T. Shanableh, and K. Assaleh, Video-based feature extraction techniques for isolated arabic sign language recognition. 9th International Symposium on Signal Processing and Its Applications (2007). doi: 10.1109/ISSPA.2007.4555408.

[31] T. Shanableh, and K. Assaleh, Two tier feature extractions for recognition of isolated Arabic sign language using Fisher's Linear Discriminants. IEEE International Conference on Acoustics, Speech and Signal Processing (2007). doi: 10.1109/ICASSP.2007.366282 .

[32] T. Shanableh, and K. Assaleh, User-independent recognition of Arabic sign language for facilitating communication with the deaf community. Digit. Signal Process., Rev. J. (2011). pp. 535-542.

[33] R. Naoum, H. H. Owaied, and S. Joudeh, Development of a new Arabic sign language recognition using k-nearest neighbor algorithm. Journal of Emerging Trends Computing and Information Sciences (2012). pp. 1173-1178.

[34] F. F. Al Mashagba, E. F. Al Mashagba, and M. O. Nassar, Automatic isolated-word Arabic sign language recognition system based on time delay neural networks: New

improvements. Journal of Theoretical and Applied Information Technology (2013). pp. 42-47.

[35] Y. El-Sonbaty, and M. A. Ismail, Matching Occluded Objects invariant to Rotations, Translation, Reflections and Scale changes. 13th Scandinavian Conf. On Image Analysis: Lecture Notes in Computer Science (2003). pp. 836-843.

[36] O. Al-Jarrah, and A. Halawani, Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artifical Intelligence (2001). pp. 117-138.

[37] T. Shanableh, K. Assaleh, and M. Al-Rousan, Spatio-temporal feature extraction techniques for isolated gesture recognition in Arabic Sign Language. IEEE Transactions on Systems, Man, Cybernetics, Part B: Cybenetics (2007)., doi: 10.1109/TSMCB.2006.889630.

[38] E. E. Hemayed, and A. S. Hassanien, Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice). in Proceedings of International Computer Engineering Conference (ICENCO) (2010). doi: 10.1109/ICENCO.2010.5720438.

[39] N. El-Bendary, et al: ArSLAT: Arabic sign language alphabets translator. International conference on Computer Information Systems and Industrial Management Applications (CISIM) (2010). doi: 10.1109/CISIM.2010.5643519.

[40] K. Assaleh, et al: Continuous Arabic sign language recognition in user dependent mode. Journal of Intelligent Learning Systems and Applications (2010), pp. 19-27.

[41] M. M. Zaki, and S. I. Shaheen, Sign language recognition using a combination of new vision based features. Pattern Recognition Letters (2011). pp. 572-577.

[42] A. A. A. Youssif, A. E. Aboutabl, and H. H. Ali, Arabic sign language (ArSL) recognition system using HMM. International Journal of Advanced Computer Science and Applications (IJACSA) (2011). pp. 45-51.

[43] N. R. Albelwi, and Y. M. Alginahi, Real-time Arabic sign language (ArSL) recognition International Conference on Communications and Information Technology (2012). pp.497-501.

[44] A. S. Elons, M Abuel-Ela, and M. F. Tolba, A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition. Applied Soft Computing (2013). pp. 1646-1660.

[45] L. Cruz, Kinect and RGBD images: Challenges and applications. 25th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials, (SIBGRAPI-T) (2012). doi: 10.1109/SIBGRAPI-T.2012.13 .

[46] J. Han et al, Enhanced computer vision with microsoft kinect sensor: A review, IEEE Transactions on Cybernetics (2013). doi: 10.1109/TCYB.2013.2265378.

[47] M. Oszust, Wysocki M.: Polish sign language words recognition with kinect. The sixth international confreence on human system interactions(HSI) (2013). doi: 10.1109/HSI.2013.6577826 .

[48] S. Moreira, Almeida, F. Guimarães, and J. Arturo Ramírez, Extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors. Expert Syst. Appl. (2014), pp.7259–7271. doi:10.1016/j.eswa.2014.05.024.

[49] L. Geng et al, Combining features for Chinese sign language recognition with Kinect. 11th IEEE International Conference on,Control & Automation (ICCA) (2014). doi:10.1109/ICCA.2014.6871127.

[50] A. Agarwal, and M. Thakur, Sign Language Recognition using Microsoft Kinect. Sixth International Conference on Contemporary Computing (IC3), Noida (2013). doi:10.1109/ICCA.2014.6871127 .

[51] M. Geetha et al, A vision based dynamic gesture recognition of Indian Sign Language on Kinect based depth images. IEEE International Conference on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA) (2013). doi:10.1109/C2SPCA.2013.6749448.

[52] Z. Zafrulla et al, American sign language recognition with the kinect. ACM, Proceedings of the 13th international conference on multimodal interfaces (2011). doi: 10.1145/2070481.2070532.

[53] K. K. Biswas, and S. K. Basu, Gesture recognition using microsoft kinect®. Proceedings of the 5th International Conference on Automation, Robotics and Applications (2011). doi: 10.1109/ICARA.2011.6144864.

[54] D. Martinez, MSc Thesis - "Sign Language Translator using Microsoft Kinect XBOX 360TM," VIBOT 5., Department of Electrical Engineering and Computer Science, Computer Vision Lab, University of Tennessee (2012).

[55] Leap Motion Controller. [Online]. Available at: https://developer.leapmotion.com/features#hand-model

[56] M. Mohandes, S. Aliyu, and M. Deriche, Arabic sign langauge recognition using the leap motion. 23rd IEEE International Symposium on Industrial Electronics (ISIE) (2014). doi: 10.1109/ISIE.2014.6864742 .

[57] M. Mohandes, Automatic translation of Arabic text to Arabic sign language. Journal on Artificial Intelligence and Machine Learning (2006). pp. 15-19.

[58] A. Almohimeed, M. Wald, and R. I. Damper, Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community. EMNLP: The Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT) (2011).

[59] A. M. Almasoud, and H. S. Al-Khalifa, A proposed semantic machine translation system for translating Arabic text to Arabic sign language. Proceedings of the 2nd Kuwait Conference on E-Services and E-Systems (2011). doi: 10.1145/2107556.2107579.

[60] H. Al-Dosri, N. Alawfi, and Y. Alginahi, Arabic sign language easy communicate ArSLEC. Proc. Int. Conf. Comput. Inf. Technol. (2012).

[61] N. Aouiti, Towards an automatic translation from Arabic text to sign language. Fourth International Conference on Information and Communication Technology and Accessibility (ICTA) (2013). doi: 10.1109/ICTA.2013.6815320.

[62] A. Samir, and M. Aboul-Ela, Error detection and correction approach for Arabic sign language recognition. Proceedings of 7th International Conference on Computer Engineering and Systems (2012). doi: 10.1109/ICCES.2012.6408496.

[63] A. M. Almasoud, and H. S. Al-Khalifa, Semsignwriting; A proposed semantic system for Arabic text-to-signwriting translation. Journal of

Software Engineering Applications (2012). pp. 604-612.

[64] F. Al Ameiri, M. J. Zamerly, and M. Al Marzouqi, Mobile Arabic sign language. International Conference for Internet Technology and Secured Transactions (ICITST) (2011).

[65] H. S. Al-Khalifa, Introducing Arabic sign language for mobile phones. Computers Helping People With Special Needs (2010), pp. 213-220.

[66] S. M. Shoheib, H. K. Elminir, and A. M. Riad, SignsWorld Atlas; a benchmark Arabic Sign Lanugage database. Journal of King Saud University-Computer and Information Sciences (2015) pp. 68-76.

[67] Y. El-Sonabty, M. Hamza and G. Basily, Compressing sets of similar medical images using multilevel centroid technique. In Proceedings of Digital Image Computing: Techniques and Applications (2003).

[68] Y. El-Sonbaty, M. A. Ismail, and E. A. El-Kwae, New algorithm for matching 2D objects. In Electronic Imaging. International society of optics and phonetics (2002), pp. 340-347.

[69] N. Sarhan, Y. El-Sonbaty and S. Youssef, HMM-Based Arabic sign language recognition using Kinect. The 10th Int. Conf. on digital Information Management (ICDIM 2015), pp. 169-174.

[70] M. Fraiwan , N. Khasawneh, H. Ershedat, I. Al-Alali, and H. Al-Kofahi, A Kinect-based system for Arabic sign language to speech translation. Int. Journal of Computer Applications in Technology. (2015), 52(2-3), pp. 117-26.