Supervised learning classifiers for Arabic gestures recognition using Kinect V2

Basma Hisham¹ · Alaa Hamouda¹

© Springer Nature Switzerland AG 2019

Abstract

Human–Computer Interaction (HCI) refers to the interaction between the computers and human. One of the most important applications of (HCI) is sign language recognition. Several research works aimed to interpret and translate the sign language to a spoken language to help the hear impaired persons in integrating with communities. Sign language is the main way of communication for the deaf persons and hearing impaired which enable them to communicate with their societies and between each other's. According to the World Health Organization, there are 466 million hearing loss people (i.e. 5% of the world population), 432 million or 93% of them are adults and 34 million or 17% of them are children. The hearing-impaired persons often need to the same level of mental capability as the normal persons. The main problem is the most of the hearing persons who cannot understand the sign language and most of hearing impaired cannot read or write our spoken language, this represents a barrier between the deaf persons and their societies so developing an automatic sign language recognition system is very necessary. This research introduces dynamic Arabic Sign Language recognition system using Microsoft Kinect. The recognition depends on using two machine learning algorithms (a) Decision Tree and (b) Bayesian Network then applied Ada-Boosting technique to enhance the recognition of the system, we compared the results with two direct matching techniques: (a) Dynamic Time Wrapping and Hidden Markov Model the system was applied on 42 Arabic gestures related to medical field. The experimental results showed that the proposed system recognition rate reached 91.18% for Decision Tree classifier, 92.50% for Bayesian classifier and 93.7% after applying Ada-Boosting.

Keywords HCI · Sign language · Static gestures · Microsoft Kinect · Decision tree · Bayesian network · DTW · HMM · Ada-boosting

1 Introduction

Sign Language is “The language that uses the manual representation and the motion and orientation of body parts such as hands, arms, wrist ...etc. in addition to the facial expression in order to represents the signer's concepts and thoughts” [1]. It is the most popular way of communication in the deaf and hearing-impaired communities also it is the natural way of communication between them and normal persons. There is a lake of awareness with the deaf societies and this led to hardness in learning and academic process, difficulties in integration and communication

with others especially the hearing persons also limitation in using the services. The hearing-impaired persons often needs to the same level of mental capability as the normal persons. The main problem is the most of the hearing persons cannot understand the sign language and most of hearing impaired cannot read or write our spoken language so this represent a barrier between the deaf persons and their societies, so that building a robust automatic sign language recognition system became necessity in order to integrate all citizens of the community with each other's regardless of their abilities [2]. According to the World Health Organization (WHO), there are 466 million

✉ Basma Hisham, basmahisham.2015@gmail.com | ¹Computer Engineering Department, Al-Azhar University, Cairo, Egypt.

Fig. 1 Deaf statistics over the years

HEARING-IMPAIRED STATISTIC IN EGYPT

Fig. 2 Hearing-impaired statistic in Egypt

hearing loss people (i.e. 5% of the world population), 432 million or 93% of them are adults and 34 million or 17% of them are children. The report expected that by the year 2050, the deaf people will reach 90 million person and this is a big number [3]. About 200 hundred million of deaf people are hearing loss in twenty-two countries around Middle East and North Africa. In Egypt, according to the last report of the “Central Agency for Public Mobilization and Statistics”, there is increasing in number of hearing-impaired persons as in Fig. 1. Nowadays, there are 7 million people suffer from hearing-impaired i.e. 10%

of the population as in Fig. 2, this sector of the community suffers from isolation, low level of education and lack of employment opportunities so that 43% of the deaf people in Egypt are unemployed and 60% of them are illiterate. The sign languages are not standard they are different as spoken language even the Arabic sign language, it differs from one country to another and varies according to dialects. This research is concerned with the Egyptian Arabic sign language which is most generally comprehended by Arabs.

From the literature we can categorized the main approaches of the sign language recognition as in Fig. 3.

Several systems were proposed in the vision based, Akmeliawati et al. [4] proposed a vision-based system for Malaysian Sign Language short vocabulary SL, they used (Video camera) they used ANN (Artificial Neural Network) for classification, they reached 90% accuracy. Hieu and Nitsuwat [5] developed a system for (SLR) they used dataset containing 19 gestures, it contains 4,121,022 colored image samples, the accuracy reached 94%, the used a simple skin color filter segmentation method which is a popular method for segmentation in vision-based approach. Adithya et al. [6] presented a model for recognizing Indian sign language based on image approach using web camera under the lighting and environmental conditions, they converted the captured images into YCbCr color space. The model depended on Discrete Fourier Transform (DFT) in features extraction and on artificial neural network (ANN) in classification phase, the back ground was restricted to be black, the classified signs were static and they achieving 91% accuracy. Nguyen et al. [7] introduced a system for Vietnamese language recognition depending on the skin color pixels static gestures of Vietnamese language using the neural network classifier. For Arabic sign language, Assaleh and Al-Rousan [8] used the vision approach for recognizing 30 Arabic alphabets, the acquired RGB images

Fig. 3 Sign language recognition categories

Table 1 Kinect-based sign language recognition systems

References	Data set and language	Methodology	Recognition accuracy (%)
Almeida et al. [12]	34 Brazilian pre-segmented words	SVM classifier (support vector machine)	80
Elleuch and Wali [13]	5 American Pre-segmented words	Multiclass SVM classifier (support vector machine)	96.8
Geng et al. [14]	20 Chinese words		69.32
Halim et al. [15]	20 Pakistanina pre-segmented words	DTW (dynamic time wrapping)	87
Dong et al. [16]	26 ASL alphabets	Random forest (RF) classifier	90
Lahiani and Elleuch [17]	10 American Sign Language	SVM classifier (support vector machine)	90
Lee et al. [18]	24 hand gestures of Taiwanese Sign Language.	HMM (Hidden Markove Model) and SVM (support vector machine)	85.14
Gkigkelos et al. [19]	15 Greek pre-segmented words	KNN (K-nearest neighbour)	99

were transformed to hue-saturation-intensity (HSI) then the features were extracted, the polynomial classifier was compared with the neuro-fuzzy inference systems and the accuracy of the latter was better and gave 93.41%. Albelwi developed [9] ArSL recognition system for alphabets and applied Haar-like classifier for hand tracking, then the skin detection is performed using HSV color space, the KNN classifier was used and achieved 90%. Ahmed and Aly [10] proposed Arabic sign language recognition system based on HMM (Hidden Markov Model), they used local binary patterns (LBP) and principal component analysis (PCA), the system was tested on 23 isolated ArSL words and achieved 99.9% in signer-dependent mode. Al Mashagba et al. [11] recognized 40 Arabic isolated words which were collected from 10 different signers. The research depended on the vision-based approach, they used the multivariate Gaussian Mixture Model (GMM) based on the characteristics of the well-known Hue Saturation Lightness Model (HIS) in order to determine the colors within the video frame, they achieved accuracy rate of 70.0% in testing set. Table 1 represents some of SLR systems which depended on Microsoft Kinect.

The sensor-based approach uses some sensors and devices in order to get the captured motion, position and velocity, from the literature we can summarize the used sensors as [20]: (a) (IMU) Inertial measurement unit which is used for recording the body position and velocity, (b) Electromyography (EMG) it is electrodiagnostic medicine device which is used for measuring and recording the human muscle's pulse in order to detect the body parts activity, (c) Radar and Wi-Fi, it uses the radio frequency signals in order to detect the human activities and body motion, (d) Other approaches depending on data gloves which are fabricated from flex sensors or including electromagnetics or ultrasonic techniques, the using of gloves is the common and popular used approach in sensor-based approach. For non-Arabic systems, Raut proposed a recognition system for 10 isolated signs using

sensor-based approach, the gloves were designed from flex sensors and accelerometer [21]. Rajamohan uses data gloves which were fabricated from 5 flex sensors and one accelerometer added to some tactile sensors, they were used to recognize 13 static American alphabets, the accuracy reached 90% [22]. Ahmad and Muhammad developed a recognition system for Malaysian Sign Language based on sensor-based approach, they used data gloves which was fabricated from 10 tilt sensor in order to capture the fingers motion and flexion and also an accelerometer in order to track the hand motion additionally, there are Bluetooth module and microcontroller unit to send the translated signs to a mobile application, the system is able to recognize some selected Malaysian letters, numbers and isolated signs, to train and test the system they collected the data from different 4 signers, the accuracy of recognizing the letters, numbers and words were 95%, 93.3% and 78.3% respectively [23]. Kim and Kyung introduced as sign language recognition system based on sensor-based technique, the glove was designed from 3-axis accelerometer, magnetometers, and also gyroscopes, all collected data was transformed into angles, the data set consists of the American numbers from 0 to 9 and the data was collected from 5 signers, the overall accuracy of recognition was 98% [24]. For Arabic sign language Assaleh et al. [25] proposed a low-complexity word-based classification for Arabic sign language recognition system using two DG5-VHand data gloves. The recognition rates were 92.5% and 95.3% for user independent and user dependent modes respectively. Mohandes [26] proposed ArSL recognition system using Cyberglove in order to track 100 two-handed signs with 20 samples and achieved accuracy is 99.6%. Sadek et al. [27] proposed a hand gesture recognition using a smart glove which was designed from a set of sensors, the recognition was based on a statistical analysis of the hand shape of during performing the 1300 words of the Arabic sign language (ArSL). Microsoft Kinect is

a motion sensing input device by Microsoft. It relies on depth technology which became an attractive aspect in sign recognition [28], it allows users to deal with any system via a camera in which the user uses hand gestures or verbal commands to recognize objects without the need to touch the controller. It used to track standing skeleton with high-depth fidelity. It has an RGB camera, voice recognition capability and face-tracking capabilities. Several researches used Kinect in object detection and tracking, human behavior analysis and sign recognition [29]. The main contribution of this paper can be summarized as the following:

- (1) Developing a new accurate and robust automatic Arabic sign language recognition system which works on real time depending on Microsoft Kinect to overcome the limitations, complexity and constraints of the previous studies (sensor-based and vision-based) such as the problems of reliability, applicability and computer vision problems such as tracking, detection and environment complexity.
- (2) Introducing a very effective normalization process which makes the system applicable for using from any signer whatever his or her position or size,
- (3) Building a robust and accurate recognition methodology based on supervised machine learning models: (a) Random Forest method as rule-Based approach and (b) Bayesian Network as a probabilistic approach.
- (4) Enhancing the proposed methodology by applying Ada-Boosting on the results of different classifiers.
- (5) Comparing the results of the model with two direct matching algorithms DTW (Dynamic Time Wrapping) algorithm and HMM (Hidden Markov Model)

It should be mentioned that ArSL is very variant and it is not a universal. Even in one country, it differs according to different cultures and dialects, some countries like Tunis has its own unified sign language but most of the countries in Middle East have different sign languages, so in this research we focused on the Egyptian Arabic sign language. The Data collection was one of the difficulties because the lackness of the Arabic data set in the area of sign language, so we built our data set from scratch depending on collecting the signed data from group of native persons for the training and testing purposes and it will be available for the Arabs research community. The structure of this paper is organized as follows. Section 2 presents the related work of the sign language recognition. The proposed system is presented in Sect. 3. The experiments setup and results are presented in Sect. 4. Section 5 contains conclusion and future work.

2 Related work

In this section we focused on the Arabic Sign Language recognition systems using different approach. Several works introduced the recognition for static Arabic Alphabets and concerned with the hand gesture recognition. Al-Jarrah and Al-Omari developed automated translation system of alphabets gestures in the Arabic Sign Language (ASL) that does not use visual markings. The images of not covered hands are processed to get a set of features that are translation, rotation, and scaling invariant. A system accuracy of 97.5% was achieved on a database of 30 Arabic alphabet signs, the paper concerned with static gestures with no regard to the body or hand motion [30]. Maraqa and Abu-Zaiter [31] introduced the use of neural networks in human hand gesture recognition for static images and dynamic gestures. They presented the use of feed-forward and recurrent neural networks and recurrent neural networks along with its different architectures. Then, they tested the proposed system, a database of 900 samples, containing 30 gestures performed by two persons with colored gloves that used in their experiments. The accuracy rate for the recognition of the static gestures reaches 95%. El-Bendary et al. [32] developed a sign language recognition system for the Arabic alphabets with an accuracy of 91.3%. The proposed Alphabets Translator (ArSLAT) does not rely on using any gloves or visual markings to complete the recognition task. ArSLAT deals with images of bare hands, which allows the signer to interact with the system in a natural way. Extracted features from a video of signs are the input to the system and the output is recognized sign as a text. The proposed ArSLAT system composed of five steps; pre-processing, detect the best-frame, detect the category, Extract the features, and finally classification. The used extracted features are translation, scale, and rotation invariant to make the system flexible. Hemayed and Hassanien [33] introduced a recognition technique for the hand gestures which represent the Arabic sign language alphabet and perform sign to voice conversion in order to enable Arabian deaf people to communicate with their societies. The proposed system focused on static and simple moving gestures. Principal Component Analysis algorithm is applied to the extracted edges that form the predefined feature vectors for signs library. The Euclidean distance is used to measure the similarity between the signs' feature, the nearest vectors sign is selected, and the corresponding sound clip is played. They applied the system to more than 150 signs and gestures with accuracy near to 97% at real time test for three different users. Naoum et al. [31, 34] developed an image-based system for sign language alphabets with an accuracy of 50%, 75%, 65% and 80% for bare hand, hand

Table 2 Arabic alphabets recognition systems

References	Data set	Classifier	Accuracy (%)
Nashwa and Zawbaa [37]	15 Arabic alphabets	Minimum distance classifier (MDC)	91.3
Marqa and Abu-Zaitar [31]	30 Arabic alphabets	Elman recurrent network	95.11
Abdo et al. [38]	29 Arabic alphabets	Hidden markov model (HMM).	100

Table 3 Arabic pre-segmented recognition systems

References	Data set	Classifier	Accuracy (%)
Youssif and Aboutabl [41]	20 Isolated Arabic sign	HMM	82.2
Mohandes [26]	120 Isolated Arabic sign	SVM	70
Hassan and Abolarinwa [42]	10 Isolated words	Artificial neural network (ANN)	90.5

with a red glove, hand with a black glove, and hand with a white glove respectively. The proposed system starts by finding histograms of the images. Extracted profiles from histograms are then used as input to the K-nearest Neighbor classifier. The algorithm is designed to work as a first level of recognition upon a series of steps to convert the captured character images into actual spelling. Elons et al. [35] proposed an ArSLR system based on neural networks that able to compensate for lighting non-homogeneity and background brightness. The proposed system showed stability under geometrical transforms, lighting conditions and bright background, achieving a recognition accuracy of 90%. The “Continuity Factor” is defined and considered as a weight factor of the current pulse in signature generation process. This factor measures the simultaneous firing strength for connected pixels. Aliyu et al. [36] proposed a new method for Arabic sign language recognition to detect and track at least one hand and at least one finger of the at least one hand from at least two different locations in a room by two different sensors; generating a 3-dimensional (3D) interaction space. They extracted 3D features associated with the one detected and tracked hand or both hands and at least one detected and tracked finger then they analyzed a relevance metric related to the extracted 3D features. They applied a classifier integrated with two different sensors—Leap Motion Controllers (LMC) and Microsoft Kinect-and 28 Arabic alphabet signs are performed in the interaction space. Table 2 listed some of Arabic alphabets recognition systems.

Other Arabic sign language recognition systems concerned with isolated signs recognition such as ElBadawy et al. [39] proposed an integrated system used a hybrid types of sensors to capture all sign features. They customized Leap motion to capture hands with fingers movements. Two digital cameras are used to capture face features and body movement. The system performed 95% recognition accuracy for a dataset of 20 dynamic signs due to the additional modules for facial expressions

recognition and body movement recognition. Aliyu et al. [36] developed a Kinect based system for Arabic sign language recognition system. The developed system was tested with 20 signs from the Arabic language dictionary. Video samples of both true color images and depth images were collected from volunteer user. Linear Discriminant analysis was used for features reduction and sign classification. Furthermore, fusion from RGB and depth sensor was carried at feature and decision level performed an overall accuracy of 99.8%. Almasre et al. [40] proposed a model to recognize the hand gestures of Arabic Sign Language (ArSL) words using two depth sensors. They examined 143 signs gestured by 10 users for 5 ArSL words. The sensors captured depth images of the upper human body, from which 235 angles (features) were extracted for each joint and between each pair of bones. The dataset was divided into a training set (109 observations) and a testing set (34 observations). They used support vector machine (SVM) classifier with different parameters on the gestured signs’ set in order to proceed four SVM models, with linear kernel (SVMLD and SVMLT) and radial kernel (SVMRD and SVMRT) functions. The accuracy of the words in the training set for the SVMLD, SVMLT, SVMRD, and SVMRT models was 88.92%, 88.92%, 90.88%, and 90.884%, respectively. The accuracy of the testing set for SVMLD, SVMLT, SVMRD, and SVMRT was 97.059%, 97.059%, 94.118%, and 97.059%, respectively. Table 3 listed some of other Arabic Pre-segmented Recognition Systems.

3 The proposed system

In this section we will discuss the structure of our proposed model for Arabic sign language recognition using Microsoft Kinect. We describe the various phases of our system from the capturing the data using Kinect until the gesture is recognized. We concerned with the features extraction and selection phase because we believe

Fig. 4 System block diagram

Fig. 5 Kinect main components

it is a very crucial phase in the recognition system. The structure of the recognition system is shown in Fig. 4. The main steps of our model can be concluded in the following: (1) Data acquisition phase which includes capturing the signer's sign and obtaining the depth image which is corresponding to this sign, (2) Preprocessing phase which includes mainly two processes firstly the joints selection and secondly the normalization process which applies a set of normalization rules to overcome the variation of the users' positions and users' sizes (3) Applying machine learning classifiers Random Forest (RF) and Naive Bayes Classifier (NBC) classifiers to recognize the captured signs then, (4) Applying Ada-Boosting on the classifiers in order to combine the results of the used classifiers to enhance the model, then (5) Applying the majority voting to select the most frequent classified class and output the sign, finally (6) Applying the DTW (Dynamic Time Warping) and HMM (Hidden Markov Model) instead of the classifiers in order to compare the results of them with the results of the machine learning approach.

3.1 Data acquisition

Kinect is a device which was developed by Microsoft on 4th of November 2010 [43], the aim of Kinect is to capture the voice and motion. The first version was Xbox 360 and it was used for video game. On 2011, Microsoft developed Kinect

for windows as a Beta version and Kinect SDK for windows was released to make Xbox 360 be used on windows 7, it can be programmed using Visual Basic.NET, C# and C++, on 2014, Microsoft released Kinect V2, it includes new features such as depth-sensing, providing skeletal tracking. The main components of Kinect are: (a) RGB camera, (b) Infrared emitter (IR), (c) multi-array microphone [44] as in Fig. 5.

The RGB camera captures a real time video for the signer, the infrared emitter produces an infrared light beams, then the depth sensor reads the reflected IR (Infra-Red) beams which are reflected again to the sensor, finally the reflected beams are converted into depth data which represents the measured distance between the signer and the sensor, Kinect used a Randomization Decision Forest classifier in order to detect the body joints position, finally Kinect uses fast, simple and effective method depending on the mean shift algorithm in order to transform the body joints positions which were detected into a skeleton as in Fig. 6.

Kinect SDK can detect the motion that is performed by the signer's skeleton and can detect the 25 skeleton joints' position as in Fig. 7 and return them as a depth information with rate of 30 frames/s.

3.2 Preprocessing

This phase includes mainly two process: (a) Joints selection and (b) Normalization.

Fig. 6 Data skeletonization

Fig. 7 Kinect Skeleton joints

3.2.1 Joint selection

Kinect's SDK performs a set of algorithms in order to manipulate the captured 3D data and infer the position in (x, y, z) coordinates, after studying the selected signs we found that only 10 joints from the 25 tracked joints are used in performing the signs, these joints are: Hand (Left and Right), Shoulder (Left and Right), Elbow (Left and Right), Wrist (Left and Right), Spin Mid and Head Center, so that there is no meaning to track other joints as the feet or knee because they will be static during performing the sign. Figure 8 represents the joints of the upper part of human body.

3.2.2 Joints' normalization

The variation in the user's position and user size can lead to recognition confusion because of any variation in depth can cause a considerable variation of the X and Y values.

Fig. 8 Points of the upper human body joint

Fig. 9 Position normalization

(a) Normalize User Position

To normalize the user's position which may be varied as in Fig. 9, we selected a reference point which will be fixed during the sign execution, we selected the spin mid joint as a reference point then the obtained distances for the captured joints will be scaled by subtracting their coordinates (x, y, z) from the coordinate of the spin mid joint, so if the signer stands in anywhere, it will be scaled and the system will not be conflicted.

All the extracted features are transferred to the spherical coordinates rather than the Cartesian coordinates X, Y, and Z, the SpinMid Joint is considerable SpinMid Joint as the origin. The spherical coordinate has three-dimensional space: the radial distance r for any point which is measured with respect to a fixed point or the origin, the (ϕ) which is the angle

between the positive x-axis and the line drawn by $r, (\theta)$ which is the angle that is measured between the positive z-axis and the line drawn from the origin to the point. We chose the spherical coordinates because it is more invariant to rotations and positions than Cartesian coordinates [45, 46]. The evaluation of the system has been showed that there is no effect to (θ) feature, so we depended on both r and ϕ and we calculated them using Eqs. 1 and 2 respectively.

The radial distance will be expressed (r) which defines the vector between the spin-mid joint (S_M) and the corresponding joint, (θ) is the angle that describes the direction of this vector.

Given a set of joints $J = \text{Left_Hand, Right_Hand, Left_Elbow, Right_Elbow, Left_Wrist, Right_Wrist, Left_Shoulder, Right_Shoulder}$.

The set of distances $r = \{r_{\text{Left_Hand}}, r_{\text{Right_Hand}}, r_{\text{Left_Elbow}}, r_{\text{Right_Elbow}}, r_{\text{Left_Wrist}}, r_{\text{Right_Wrist}}, r_{\text{Left_Shoulder}}, r_{\text{Right_Shoulder}}\}$

The set of orientations $= \{\theta_{\text{Left_Hand}}, \theta_{\text{Right_Hand}}, \theta_{\text{Left_Elbow}}, \theta_{\text{Right_Elbow}}, \theta_{\text{Left_Wrist}}, \theta_{\text{Right_Wrist}}, \theta_{\text{Left_Shoulder}}, \theta_{\text{Right_Shoulder}}\}$

$$\sum_{i=1}^n r(i) = \sqrt{(J_x(i) - S_{M_x})^2 + (J_y(i) - S_{M_y})^2 + (J_z(i) - S_{M_z})^2} \quad (1)$$

$$\sum_{i=1}^n \theta(i) = \text{atan2}((J_y(i) - S_{M_y}), (J_x(i) - S_{M_x})) \quad (2)$$

where, $r(i)$ is a radial distance for a specific joint i , J is the joint of interest, $\theta(i)$ the polar angle for joint i , n is the number of joints from J , S_{M_x} is x coordinate of spin-mid joint, S_{M_y} is y coordinate of spin-mid joint, S_{M_z} is z coordinate of spin-mid joint.

(b) Normalize User Size

The user size may vary from one signer to another as in Fig. 10 and this may lead to conflict, so it is necessary to normalize all joints' distances that are captured. The distance between the head and spin-mid (r_{H,S_M}) is used to normalize the signer's size, this value is calculated in Eq. 3.

$$\sum_{i=1}^n r_{\text{norm}}(i) = \frac{r(i)}{r_{H,S_M}} \quad (3)$$

where, n is the number of joints from J , r_{norm} is a normalized radial distance of the joint, r_{H,S_M} is a radial distance from head center to spin-mid it can be calculated as in Eq. 4:

Fig. 10 User's size normalization

$$r_{H,S_M} = \sqrt{(J_x(i) - S_{M_x})^2 + (J_y(i) - S_{M_y})^2 + (J_z(i) - S_{M_z})^2} \quad (4)$$

3.3 Joints' angles calculation

The joint angles are very important and essential features because these angles differ from one gesture to another and therefore, they are one of the most important features that can be used in recognition and distinguishing between gestures. We have chosen a set of angles that affect the performance of the system strongly. These angles were formed from three joints, the angles are listed below:

1. Shoulder–Elbow–Wrist (Left and Right) angles which are made by both shoulder and wrist on the elbow, these angles are calculated on both sides (Left and Write) as in Fig. 11a.
2. Elbow–Shoulder–Shoulder Center (Left and Right) angles which are made by both elbow and shoulder center on the shoulders on both sides (Left and Write) as in Fig. 11b.
3. Elbow–Shoulder–Spin mid (Left and Right) angles which are made by both elbow and Spin mid on the shoulders on both sides (Left and Write) as in Fig. 11c.

3.4 Features selection

As it was mentioning the used joints are Hand (Left and Right), Shoulder (Left and Right), Elbow (Left and Right),

Fig. 11 Joint angles **Fig. 12** Arabic dynamic gestures majority voting

Wrist (Left and Right), also Spin Mid and Head Center because they are the joints that are used in performing the sign and other remaining joints are static. The features of interest were selected by running an automatic features selection algorithm namely “Filter features selection”, the resultant features were as the following:

1. The joints' positions in spherical coordinates (Radial distance (r), and orientation ($\emptyset$) (16 Features)

$\{Hand_{Left_r}, Hand_{Right_r}, Wrist_{Left_r}, Wrist_{Right_r}, Elbow_{Left_r}, Elbow_{Right_r},$
 $Shoulder_{Left_r}, Shoulder_{Right_r}, Hand_{Left_{\emptyset}}, Hand_{Right_{\emptyset}},$
 $Wrist_{Left_{\emptyset}}, Wrist_{Right_{\emptyset}}, Elbow_{Left_{\emptyset}}, Elbow_{Right_{\emptyset}}, Shoulder_{Left_{\emptyset}}, Shoulder_{Right_{\emptyset}}\}$

2. The difference in distance between the joints' positions (16 Features)

$\{Hand_{left} - Elbow_{Right}, Hand_{left} - Hand_{Right}, Hand_{left} - Hand_{Right}, Hand_{left} - Wrist_{Right}, Hand_{left} - Shoulder_{Right},$
 $Elbow_{left} - Elbow_{Right}, Elbow_{left} - Hand_{Right}, Elbow_{left} - Wrist_{Right}, Elbow_{left} - Shoulder_{Right}, Wrist_{left} - Wrist_{Right},$
 $Wrist_{left} - Hand_{Right}, Wrist_{left} - Elbow_{Right}, Wrist_{left} - Shoulder_{Right}, Shoulder_{left} - Shoulder_{Right}, Shoulder_{left} - Hand_{Right},$
 $Shoulder_{left} - Elbow_{Right}, Shoulder_{left} - Wrist_{Right}\}$

3. Set of Angles between joints (6 Angles)

$\{(Elbow_{Right} - Shoulder_{Right} - Shoulder_{center})\text{Angle}, (Elbow_{Left} - Shoulder_{Left} - Shoulder_{center})\text{Angle},$
 $(Shoulder_{Right} - Elbow_{Right} - Wrist)\text{Angle}, (Shoulder_{Left} - Elbow_{Left} - Wrist)\text{Angle},$
 $(Elbow_{Right} - Shoulder_{Right} - Spin_{Mid})\text{Angle}, (Elbow_{Left} - Shoulder_{Left} - Spin_{Mid})\text{Angle}\}$

3.5 Classification

The final step of our model was the classification. In order to achieve the recognition, goal the collected frames should be coupled with an effective features extraction followed by highly efficient classification. For our proposed model we used two classifiers (a) Naive Bayes classifier and, (b) Random Forest classifier. The result finally is selected by majority voting i.e. the classification with the

most frequent frames assigned to it because of the dynamic nature of gestures. For example, if the

Fig. 13 Ada-boosting structure **Fig. 14** Random forest structure

classification results of the test sequence were as in Fig. 12, then the major index will be index "1" so the translated sign will be sign with index = 1 (تقيؤ/Vomiting). Finally, we applied the Ada-Boosting to enhance the results of the classification. It is an ensemble classifier that was firstly introduced by Freund and Schapire in 1996. It is constructed from multiple classifiers or set of weak classifiers, the single classifier can act poorly but the results of multiple classifiers are expected to be enhanced and more accurate. We applied the Ada Boosting method on two classifiers NBC (Naive Bayes) classifier, RF (Random Forest) classifier as in Fig. 13. The main process of the Ada Boosting starts by giving equal weights to each observation in data set, if the prediction of the first classifier is incorrect, then it gives the highest weight for the observation with incorrect prediction, then it gives the highest weight for the observation with incorrect prediction.

3.6 Random forest classifier

Random Forest Algorithm is used to overcome the issues of Decision Tree, and it is an ensemble learning method which can be used in several ways as: classification and

regression. The main process of random forest is constructing the set of decision trees as in Fig. 14 during the training phase and finally output the class in the case of classification or the prediction in the case of regression [47].

3.7 Naive Bayes classifier (NBC)

NBC is the a simple and powerful probabilistic classifier based on Bayesian theorem $P(A|C_1) \dots P(A|C_n)$ with the main assumption of the features independency. It allows to predict the happening evidence according to other evidences and their probabilities. For Naïve Bayes, given the instances of the problem and getting the n Features vector, then it assigns the probabilities to the instances $p(C_k|x_1, \dots, x_n)$, for each K possible or classes C_k [48] as in Fig. 15.

3.8 DTW (dynamic time warping)

One of the most used algorithms in dynamic gestures recognition is DTW (Dynamic Time Wrapping) which is a pattern recognition algorithm. It is used in several applications such as speech recognition, gestures recognition,

Fig. 15 Naive Bayes classifier process **Fig. 16** Two aligned time series sequences **a** LH. D feature for a sample of sign "نزيف" or "Bleeding" stored in training set **b** LH. D feature for a sample of sign "نزيف" or "Bleeding" stored in testing set **c** LH. D feature for a sample of sign "اشعة" or "X-Ray" stored in testing set**Fig. 17** LH. D (Left hand-distance) feature over samples of two different signs

signal processing and surveillance. DTW finds the optimal alignment between two time series sequences which are varying in their speed or their time and also have different lengths. The main idea of DTW is wrapping the two sequences non-linearly in time by measuring the distance between them, Fig. 16 represents the process of alignment for two time-independent sequences. Given two time series $X = (x_1, x_2, x_4, \dots, x_n)$ with length $n \in \mathbb{N}$ and $Y = (y_1, y_2, y_4, \dots, y_n)$ with length $m \in \mathbb{N}$ and let F is a feature space where $X_n, Y_m \in F$. DTW will analyze the sequences in order to find the similarities between them and finally find optimal alignment $O(MN)$ [46].

Figure 17 represents the difference between the two signs "نزيف" or "Bleeding" with $ID = 39$ and "اشعة" or "X-Ray" with $ID = 2$, the graphs represents the orientation of left-hand distance (LH. D), X-axis represents the number of

frames and Y-axis represents the coded values corresponded to each frame. The first two graph represent the similarities between two samples for the same sign which performed from two different users at different time and the third graph represents a different sign and it is clear that it is completely different from them.

Figure 18 represents the original signal and warped signal for the left hand-distance feature for the sign "نزيف" or "Bleeding", also Fig. 19 represents the DTW graph between the original signal and aligned sign

Figures 20, 21 represents the final step of DTW, if we have a data set of 3 stored signs, and a new sign is performed, the DTW algorithm computes the distance between the captured sign and each sign in the training set. Finally, the algorithm matches the sign with the stored sign which is corresponding to the least distance.

Fig. 18 Original signal and warped signal for left-hand distance (LH. D) over a sample for Sign_ID (39) “نزيف” or “Bleeding”

Fig. 19 DTW graph between the original signal and aligned sign for the sign “نزيف” or “Bleeding”

3.9 HMM (hidden Markov model)

It is a stochastic model which works in time domain, and it is applied successfully in speech recognition and gesture recognition applications. Basically, the Markov model has two parts: (a) set of finite states and (b) the probabilities of the emissions from one state to another. The “hidden”

expression comes from that HMM model AND a set of hidden states are constructed from the observation sequences. Mainly, there are two phases for the HMM model training and testing. Regarding the training phase, Fig. 22 represents building the HMM states for one gestures “Colic/مغص” as an example while in the testing phase,

Fig. 20 DTW classification example **Fig. 21** HMM state model for gesture

Fig. 23 represents the work flow of the system when using the HMM in dynamic gestures recognition.

4 Experimental results

For the purpose of studying the performance of the proposed system, we developed a software application based on the C# Language to interface with Kinect sensor and we used Microsoft Kinect SDK, Weka is used for building the classifier model. The overall workflow of the proposed system is composed by 5 steps as presented in Fig. 23. A prototype program was developed in C# for data collection as in Fig. 24. The program was designed with GUI which allows the user to start recording Kinect frames and recognizing the performed sign.

4.1 Arabic signs data set

For the purpose of studying the performance of the proposed system, we recorded 42 different gestures in medical field which are listed in Table 4 using Microsoft Kinect V2. The data collected from two different volunteers in different position and with different sizes.

4.2 Training and testing

For training phase, we collected 20 samples from different 5 signers for each sign as a training set and 10 samples from different 2 signers as a testing set. Thus, making it a total of 1260 samples which we divided into 840 samples for train set and 420 samples for test set. Figures 25, 26, 27 and 28 represent examples of captures signs (الاسهال/Dysentery), (المغص/Colic), (الصداع/A headache) and (الربو/Allergy).

Fig. 22 HMM model for dynamic gestures recognition

Fig. 23 Proposed model workflow

Two-Lungs) respectively during performing them. All collected signs are dynamic (i.e.) multiple joints are moving like hand, wrist, elbow and shoulder. For each sign we recorded a sequence of skeleton data consisted of 20 joint positions per frame, which are formed from x, y, and depth coordinates). Each incoming frame are preprocessed, the joints of interest are selected and normalized relative to SpinMid joint as in Sect. 3.3. Each sign's sequence contains on average (100 or 170) frames, making for around 36,456 frames in total for training set and 33,995 frames for testing set. Then the feature vector is formed by applying our mentioned technique and appending the extracted features which are recorded for each frame. This is done for both the training as well as the test dataset. We applied

two different classifiers in the classification phase (NBC and RF) for comparing the achieved accuracy of them. RF classifier achieved the best results in classification of separated frames.

4.3 Accuracy measures

There are two types of recognition: (1) Model-Based matching such as machine learning techniques (Classifiers as NBC, ANN, KNN, SVM, Decision Tree, Random Forest...etc.) and (2) Direct matching techniques such as DTW (Dynamic Time Warping), HMM (Hidden Markov Model), maximum correlation coefficient. In our research, we tested the model by both techniques firstly with a set of classifiers and compared the results of them, then the system was tested using DTW. The results of our system are divided into two parts: the accuracy of recognition and the time of response.

4.3.1 Classifiers accuracy

We have empirically calculated the accuracy which is defined as the ratio of correctly recognized frames to the total number of frames as in Eq. 5

Accuracy

$$= \frac{\text{Total number of correctly recognized instances}}{\text{Total number of instances}} \times 100\% \quad (5)$$

Fig. 24 Application GUI

Table 4 Medical dataset

Index	Arabic sign	Meaning in English	Index	Arabic sign	Meaning in English
1	اسهال	Dysentery	22	نزيف	Bleeding
2	اشعة	X-Ray	23	وفاة	Death
3	استقبال	Reception	24	طبيب عظام	Orthopedic doctor
4	رئتان	Two lungs	25	علاج طبيعي	Physical therapy
5	كبد	liver	26	حقن	Injection
6	كلى	kidneys	27	زغللة	Blurred vision
7	معدة	stomach	28	سرطان	Cancer
8	امساك	Constipation	29	جهاز قياس ضغط	Pressure measuring device
9	تحليل	analysis	30	صداع	A headache
10	تطعيم	Vaccination	31	صمم	Deafness
11	شلل	Paralysis	32	طبيب اطفال	Pediatrician
12	طبيب توليد	Obstetrician	33	طبيب انف واذن	Doctor of nose and ear
13	تقيوء	Vomiting	34	طبيب باطنة	Internist
14	تورم	Swelling	35	طبيب عام	General Doctor
15	جرح	Wound	36	كسر عظام	Broken bones
16	حامل	Pregnant	37	فيتامينات	Vitamins
17	حرارة	Fever	38	فشل كلوى	Kidney failure
18	اوردة	Veins	39	قرحة	ulcer
19	حساسية	Allergic	40	قولون	The colon
20	مغص	Colic	41	مختبر	laboratory
21	انا	I	42	اشعر	Feel

For example, if we have a signIndex_3 with 150 frames to be classified that 30 frames classified as signIndex_1, 20 frames classified as signIndex_2 and 100 frames classified

as signIndex_3 so after applying majority voting, the final predicted class will be signIndex_3. In previous example, accuracy of frames classification equals $\frac{100}{150} = 67\%$ and

Fig. 25 Gesture of (لاسه/ dysentery)

Fig. 26 Gesture of (صغم/ colic)

Fig. 27 Gesture of (عادص/ headache)

after applying majority voting, the accuracy equals 100%. We used NBC (Naïve Bayes classifier) and RF classifiers in the classification phase to compare the accuracy of classification for the separated frames, we got the accuracy of 84.11% and 86% respectively for frames classification (without majority voting) in addition to 91.18%, 92.50% (with majority voting) as it is shown in Fig. 29 It is noticed that applying the majority voting on the set of frames to get the sign meaning had enhanced the accuracy. Finally, the accuracy enhanced after using Ada-boosting to reach

93.7%. Figure 30 represents the classifiers accuracy for each sign.

4.3.2 Direct matching accuracy (DTW and HMM)

We also tested the proposed system using DTW and (HMM with 25 hidden states) algorithms and applied them on the same test data set. The overall accuracies that were achieved using The DTW and HMM reach 77.8% and 79.5%

Fig. 28 Gesture of (عاص/ headache) **Fig. 29** Classifiers accuracy comparison

respectively. Figure 31 shows the results of applying both DTW and HMM on the testing samples.

4.3.3 Classifiers response time

The measurable criteria for the system performance are a very important metrics, especially when the system works in real-time so the response time should be computed. To calculate the average time which is taken by the proposed system to recognize the performed signs, we calculate the

time of response for each samples of the testing data set and finally get the average of the calculated response time. For both classifiers (NBC and RF) and also for direct matching techniques (DTW and HMM), Fig. 32 represents the response time over the 10 samples of the test-data set for sign “اشعة” or “X-Ray”. It is clear that the recognition process with SVM classifier has the least response time. Also Fig. 33 represents the response time for the 42 signs using the different classifiers over the test data set (10 sample per sign). It is clear that the system takes the least response time when we use NBC in classification step (8.09 S) as in average, then RF (15.3 S) then DTW (17.63 S) and finally HMM (20.8 S)

4.3.4 Comparative study

Because of the absence of the comparative studies and benchmark dataset, so we introduced a visible comparison using CHALEAR dataset which is available on [49]. The ChaLearn 2013 dataset contains 20 Italian gestures which were recorded using Kinect. It contains (Both depth and RGB images for face and body, skeleton data, Joint orientation and position also audio sources) as in Fig. 34. It includes 15,000 samples from several signers,

Fig. 30 Classifiers accuracy for each sign

Fig. 31 HMM and DTW accuracies for each sign **Fig. 32** Response time for recognition of "اشعة" or "X-Ray" **Fig. 33** Average response time

they are divided to three parts: training (7754 gestures), validation (3362 gestures), and testing data (2742 gestures). An example of the ChaLearn 2013 dataset is shown in Fig. 34 and the result of our comparison is listed in Table 5.

5 Conclusion

In this paper, dynamic Arabic sign recognition system using Microsoft Kinect is proposed. The proposed system is developed and applied on medical words. Captured user's position and size are normalized to solve

Fig. 34 Samples from ChaLearn training dataset

Table 5 ChaLearn data set comparative study results

Dataset	Number of signers	Number of gestures	Accuracy [50] (2015)	Accuracy [52] (2013)	Accuracy [51] (2017)	Proposed model	Enhanced model using ada-boosting
CHALEARN 2013	27 signers	20	SVM 76% (7 joints)	DTW 59.9% Skeleton data	HMM 63.34% (4 joints)	Classification 80%	88%

the variation problem in collected frames. Feature selection method is applied on collected frames and 38 effectives were selected. We applied two techniques to recognize the captured signs: (1) Machine learning classifiers such as Naive Bayes and Decision Random Forest and (2) Direct matching technique as DTW and then the results were compared. System performance is analyzed using 42 Arabic words from medical field. The system was trained on 1260 samples and tested on 840 samples, the experimental results showed that the proposed system recognition rate reached to reach to 91.18%, 92.50% for Naive Bayes, Random Forest and the model reach 93.7% after Ada Boosting. For DTW and HMM, the accuracy reached 77.8% and 79.5% respectively. After computing the latency of the system, Naive Bayes takes the least response time step (8.09 S) as in average, then RF (15.3 S) then DTW (17.63 S) and finally HMM (20.8 S). We introduced a comparative study using CHALEARN dataset, our proposed model achieved a progress in accuracy reached 88% compared to other researches. For further research, we will enhance the accuracy of the recognition and system latency, extend the data set to incorporate more vocabularies in different fields,

enable the model to recognize the Arabic continues sentences and applied Arabic sign language grammar rules in order to make the recognized sentences have a useful meaning also make the system able to recognize the facial expressions, finally applying the Deep learning which will improve the accuracy and efficiency of sign language recognition system and also help to bring real-time response to the life and improve the quality of life of the deaf people.

Compliance with ethical standards

Conflict of interest The Author stated that there is no conflict of interest and the research has no funds also.

References

1. Kausar S, Javed MY (2011) A survey on sign language recognition. In: 2011 frontiers of information technology. IEEE, pp 95–98
2. Mohandes M, Deriche M, Liu J (2014) Image-based and sensor-based approaches to Arabic sign language recognition. IEEE Trans Hum Mac Syst 44(4):551–557

3. Charayaphan C, Marble AE (1992) Image processing system for interpreting motion in American Sign Language. *J Biomed Eng* 14(5):419–425
4. Akmeliawati R, Leen Ooi MP, Kuang YC (2007) Real-time Malaysian Sign Language translation using colour segmentation and neural network. In: Instrumentation and measurement technology conference (IMTC2007), Warsaw, Poland, pp 1–6
5. Hieu DV, Nitsuwat S (2008) Image preprocessing and trajectory feature extraction based on hidden markov models for sign language recognition. In: 2008 ninth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing, Phuket, Thailand, pp 501–506
6. Adithya V, Vinod PR, Gopalakrishnan U (2013) Artificial neural network-based method for Indian sign language recognition. In: 2013 IEEE conference on information and communication technologies (ICT), Thuckalay, Tamil Nadu, India, Indiavol, pp 1080
7. Nguyen T-N, Huynh H-H, Meunier J (2013) Static hand gesture recognition using artificial neural network. *J Image Graph* 1(1):34–38
8. Assaleh K, Al-Rousan M (2005) Recognition of Arabic sign language alphabet using polynomial classifiers. *EURASIP J Adv Signal Process* 2005:507614. <https://doi.org/10.1155/ASP.2005.2136>
9. Albelwi NR, Alginahi YM (2012) Real-time Arabic sign language (Arsl) recognition. In: International conference on communications and information technology, Madinah, Saudi Arabia
10. Ahmed AA, Aly S (2014) Appearance-based arabic sign language recognition using hidden markov models. In: 2014 international conference on engineering and technology (ICET). IEEE, Cairo, pp 1–6
11. Al Mashagba FF, Al Mashagba EF, Nassar MO (2013) Automatic isolated-word Arabic sign language recognition system based on time delay neural networks: new improvements. *J Theor Appl Inf Technol* 57(1):42–47
12. Almeida SGM, Guimarães FG, Ramírez JA (2014) Feature extraction in brazilian sign language recognition based on phonological structure and using rgb-d sensors. *Expert Syst Appl* 41(16):7259–7271
13. Elleuch H, Wali A, Samet A, Alimi AM (2015) A static hand gesture recognition system for real time mobile device monitoring. In: 2015 15th international conference on intelligent systems design and applications (ISDA). IEEE, Marrakech, pp. 195–200
14. Geng L, Ma X, Xue B, Wu H, Gu J, Li Y (2014) Combining features for chinese sign language recognition with kinect. In: 11th IEEE international conference on control and automation (ICCA). IEEE, Taichung, pp 1393–1398
15. Halim Z, Abbas G (2015) A Kinect-based sign language hand gesture recognition system for hearing-and speech-impaired: a pilot study of Pakistani sign language. *Assist Technol* 27(1):34–43
16. Dong C, Leu MC, Yin Z (2015) American sign language alphabet recognition using microsoft kinect. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 44–52
17. Lahiani H, Elleuch M, Kherallah M (2015) Real time hand gesture recognition system for android devices. In: 2015 15th international conference on intelligent systems design and applications (ISDA). IEEE, Marrakech, pp 591–596
18. Lee GC, Yeh F-H, Hsiao Y-H (2016) Kinect-based Taiwanese sign-language recognition system. *Multimed Tools Appl* 75(1):261–279
19. Gkigkelos N, Goumopoulos C (2017) Greek sign language vocabulary recognition using Kinect. In: Proceedings of the 21st Pan-Hellenic conference on informatics. ACM, Larisa
20. Cheok MJ, Omar Z, Jaward MH (2019) A review of hand gesture and sign language recognition techniques. *Int J Mach Learn Cybern* 10(1):131–153
21. Shukor AZ et al (2015) A new data glove approach for Malaysian sign language detection. *Procedia Comput Sci* 76:60–67
22. Kim K-W, Lee M-S, Soon B-R, Ryu M-H, Kim J-N (2016) Recognition of sign language with an inertial sensor-based data glove. *Technol Health Care* 24(s1):S223–S230
23. Abdo MZ, Hamdy AM, Salem SAE-r, Saad E-sM (2014) Arabic sign language recognition. *Int J Comput Appl* 89(20):0975–8887
24. Kau L-J, Su W-L, Yu P-J, Wei S-J (2015) A real-time portable sign language translation system. In: 2015 IEEE 58th international midwest symposium on circuits and systems (MWSCAS). IEEE, Fort Collins, pp 1–4
25. Assaleh K, Shanableh T, Zourob M (2012) Low complexity classification system for glove-based arabic sign language recognition. In: International conference on neural information processing. Springer, Berlin, pp 262–268
26. Mohandes MA (2013) Recognition of two-handed Arabic signs using the CyberGlove. *Arab J Sci Eng* 38(3):669–677
27. Sadek MI, Mikhael MN, Mansour HA (2017) A new approach for designing a smart glove for Arabic Sign Language Recognition system based on the statistical analysis of the sign language. In: 2017 34th national radio science conference (NRSC). IEEE, Alexandria
28. Ma C, Wang A, Chen G, Chi X (2018) Hand joints-based gesture recognition for noisy dataset using nested interval unscented Kalman filter with LSTM network. *Vis Comput* 34(6–8):1053–1063
29. Han J et al (2013) Enhanced computer vision with microsoft kinect sensor: a review. *IEEE Trans Cybern* 43(5):1318–1334
30. Al-Jarrah O, Al-Omari FA (2007) Improving gesture recognition in the Arabic sign language using texture analysis. *Appl Artif Intell* 21(1):11–33
31. Maraqa M, Abu-Zaiter R (2008) Recognition of Arabic sign language (ArSL) using recurrent neural networks. In: 2008 first international conference on the applications of digital information and web technologies (ICADIWT). IEEE, Ostrava, pp 478–481
32. El-Bendary N, Zawbaa HM, Daoud MS, Hassanien AE, Nakamatsu K (2010) Arslat: Arabic sign language alphabets translator. In: 2010 international conference on computer information systems and industrial management applications (CISIM). IEEE, Krakow, pp 590–595
33. Hemayed EE, Hassanien AS (2010) Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice). In: 2010 international computer engineering conference (ICENCO). IEEE, Giza, pp 121–127
34. Reyadh Naoum SJ, Owaied HH (2012) Development of a new arabic sign language recognition using k-nearest neighbor algorithm. *J Emerg Trends Comput Inf Sci* 3:1173–1178
35. Elons A Samir, Abull-Ela M, Tolba MF (2013) A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition. *Appl Soft Comput* 13(4):1646–1660
36. Aliyu S, Mohandes M, Deriche M, Badran S (2016) Arabie sign language recognition using the Microsoft Kinect. In: 2016 13th international multi-conference on systems, signals and devices (SSD). IEEE, Leipzig, pp 301–306
37. El-Bendary N, Zawbaa HM, Hassanien AE, Snasel V (2011) PCA-based home videos annotation system. *Int J Reason Based Intell Syst* 3(2):71–79
38. Abdo M, Hamdy A, Salem S, Saad EM (2015) Arabic alphabet and numbers sign language recognition. *Int J Adv Comput Sci Appl* 6(11):209–214

39. ElBadawy M, Elons AS, Sheded H, Tolba MF (2015) A proposed hybrid sensor architecture for arabic sign language recognition. In: *Intelligent systems' 2014*. Springer, Cham, pp 721–730
40. Almasre M, Al-Nuaim H (2017) Comparison of four SVM classifiers used with depth sensors to recognize Arabic sign language words. *Computers* 6(2):20
41. Youssif AA, Aboutabl AE, Ali HH (2011) Arabic sign language (Arsl) recognition system using hmm. *Int J Adv Comput Sci Appl* 2(11):45–51
42. Hassan ST, Abolarinwa JA, Alenoghena CO, Bala SA, David M, Farzaminia A (2017) Intelligent sign language recognition using enhanced fourier descriptor: a case of Hausa sign language. In: *2017 IEEE 2nd international conference on automatic control and intelligent systems (I2CACIS)*. IEEE, Kota Kinabalu, pp 104–109
43. Kinect for windows sdk beta launch. <http://channel9.msdn.com/Events/KinectSDK/BetaLaunch>. June 2011. Last checked 24 July, 2014
44. Eisler C (2012) Starting February 1, 2012: use the power of Kinect for windows to change the world. <http://blogs.msdn.com/b/Kinectforwindows/archive/2012/01/09/kinect-for-windows-commercial-program-announced.aspx>. January 2012. Last checked 24 July, 2014
45. Vogler C, Metaxas D (1998) ASL recognition based on a coupling between HMMs and 3D motion analysis. In: *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, pp 363–369
46. Premaratne P (2014) Effective hand gesture classification approaches. In: *Human computer interaction using hand gestures*. Springer, Singapore, pp 105–143
47. Joshi A, Monnier C, Betke M, Sclaroff S (2017) Comparing random forest approaches to segmenting and classifying gestures. *Image Vis Comput* 58:86–95
48. Pramunanto E, Sumpeno S, Legowo RS (2017) Classification of hand gesture in indonesian sign language system using Naive Bayes. In: *2017 international seminar on sensors, instrumentation, measurement and metrology (ISSIMM)*. IEEE, pp 187–191
49. <https://www.kaggle.com/c/multi-modal-gesture-recognition>
50. Escobedo-Cardenas E, Camara-Chavez G (2015) A robust gesture recognition using hand local data and skeleton trajectory. In: *2015 IEEE international conference on image processing (ICIP)*. IEEE, pp 1240–1244
51. Chen X, Koskela M (2014) Using appearance-based hand features for dynamic rgb-d gesture recognition. In: *2014 22nd international conference on pattern recognition (ICPR)*. IEEE, pp 411–416
52. Kumar P, Saini R, Roy PP, Dogra DP (2018) A position and rotation invariant framework for sign language recognition (SLR) using Kinect. *Multimed Tools Appl* 77(7):8823–8846

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.