Arabic Sign Language (ArSL) Recognition System Using HMM

Aliaa A. A.Youssif , Amal Elsayed Aboutabl, Heba Hamdy Ali

Department of Computer Science, Faculty of Computers and Information

Helwan University

Cairo, Egypt

Abstract—**Hand gestures enabling deaf people to communication during their daily lives rather than by speaking. A sign language is a language which, instead of using sound, uses visually transmitted gesture signs which simultaneously combine hand shapes, orientation and movement of the hands, arms, lip-patterns, body movements and facial expressions to express the speaker's thoughts. Recognizing and documenting Arabic sign language has only been paid attention to recently. There have been few attempts to develop recognition systems to allow deaf people to interact with the rest of society. This paper introduces an automatic Arabic sign language (ArSL) recognition system based on the Hidden Markov Models (HMMs). A large set of samples has been used to recognize 20 isolated words from the Standard Arabic sign language. The proposed system is signer-independent. Experiments are conducted using real ArSL videos taken for deaf people in different clothes and with different skin colors. Our system achieves an overall recognition rate reaching up to 82.22%.**

Keywords-Hand Gesture; Hand Tracking; Arabic Sign Language (ArSL); HMM; Hand Features; Hand Contours.

I. INTRODUCTION

Singing has always been part of human communications [1]. For millennia, deaf people have created and used signs among themselves. These signs were the only form of communication available for many deaf people.Within the variety of cultures of deaf people all over the world, signing evolved to form complete and sophisticated languages. These languages have been learned and elaborated by succeeding generations of deaf children.

Normally, there is no problem when two deaf persons communicate using their common sign language. The problem arises when a deaf person wants to communicate with a non-deaf person. Usually both will be dissatisfaction in a very short time.

In this section we focus our discussion of the efforts made by researchers on sign language gesture recognition in general and on Arabic sign language (ArSL) in particular. Sign language recognition systems can be further classified into signer-dependent and signer-independent. Also one may classify. Sign language recognition systems are either glove-based which relies on electromechanical devices for data collection, or none glove-based if free hands are used. The learning and recognition methods used in previous studies to recognize sign language include neural networks and hidden Markov models (HMMs).

Cyber gloves have been widely used in most of previous Works on sign language recognition including [1, 2, 3]. Kudos [4] reported a system using power gloves to recognize a set of 95 isolated Australian sign languages with 80% accuracy. Grobel and Assan [5] used HMM to recognize isolated signs with 91.3% accuracy out of a 262-sign vocabulary. They extracted 2D features from video recordings of signers wearing colored gloves. Colored gloves were used in [6] where HMM was employed to recognize 52 signs of German sign language with a single color video camera as input. In a similar work [7] an accuracy of 80.8% was reached in the corpus of 12 different signs and 10 subunits using the K-means clustering algorithm to get the subunits for continuous sign language recognition. Liang and Ouhyoung [8] employed the time-varying parameter threshold of hand posture to determine end-points in a stream of gesture input for continuous Taiwanese sign language recgnation with the average recognition rate of 80.4% over 250 signs. In their system HMM was employed, and data gloves were taken as input devices.

The use of cyber gloves or other means of input devices conflicts with recognizing gestures in a natural context and is very difficult to run in real time. Therefore, recently researchers presented several sign recognition systems based on computer vision techniques [9, 10, 11]. Starner et al [12] used a view-based approach for continuous American continuous sign language recognition. They used a single camera to extract two-dimensional features as the input of HMM. The accuracy of 92% or 98% was obtained when the camera was mounted on the desk or in a user's cap in recognizing the sentences with 40 different signs. However, the user must wear two colored gloves (a yellow glove for the right hand and an orange glove for the left) and sits in a chair before the camera. Vogler and Metaxas [14] used computer vision methods and interchangeably AT a flock of birds for 3D data extraction of 53 signs for American Sign Language. They, respectively, built context-dependent HMM and modeled transient movement to alleviate the effects of movement epenthesis. Experiments over 64 phonemes extracted from 53 signs showed that modeling the movement epenthesis has better performance than context-dependent HMM. The system provided overall accuracy of 95.83% is reported.

Furthermore, most of the above systems are signer dependent systems. A more convenient and efficient system is the one that allows deaf users to perform gestures naturally with no prior knowledge about the user. To the best of our knowledge there are a few published works on signer-independent systems. Vamplew and Adams [13] proposed a signer-independent system to recognize a set of 52 signs. The system used a modular architecture consisting of multiple feature-recognition neural networks and the nearest neighbor classifier to recognize isolated signs. They reported a recognition rate of 85% in the test set. Again the signer must wear cyber gloves while performing gestures. Another attempt is made by Fang et al [14] in which they used the SOFM/HMM model to recognize signer-independent CSL over the 4368 samples from 7 signers with 208 isolated signs.

ArSLs are still in their development stages. A glove-based and singer-dependant Arabic sign recognition system has been developed by M. Mohandes and S. I. Quadri, M. Deriche [15]. They used a data set of 15 samples for each of the 300 signs which were carried out by a signer wearing a pair of colored gloves (orange and yellow) achieving recognition accuracy about 88.73%. Jarrah and Halawani [3] developed a system for ArSL alphabet recognition using a collection of Adaptive Neuro-Fuzzy Inference Systems, a form of supervised learning. They used images of bare hands instead of colored gloves to permit the user to interact with the system conveniently. The used feature set comprised lengths of vectors that were selected to span the fingertips' region and training was accomplished by the use of a hybrid learning algorithm achieving recognition accuracy of 93.55%. Likewise, Assaleh and Rousan [1] extended the work in [3] by using Polynomial classifiers extracted superior results on the same dataset. Their work required the participants to wear gloves with colored tips while performing the gestures to simplify the image segmentation stage. They extracted features including the relative position and orientation of the fingertips with respect to the wrist and to each other. The resulting system achieved 93.41% recognition accuracy.

More recently, recognition of video-based isolated Arabic sign language gestures is reported by in [16] and [17]. In [16], the dataset is based on 23 gestures performed by 3 signers. The data collection phase did not impose any restrictions on clothing or background. Forward or bi-directional prediction error of the input video sign was accumulated and threshold into a single image. The still image is then transformed into the frequency domain. The feature vector that represents the gesture is then based on the frequency coefficients. Simple classification techniques such as KNN, linear and Bayesian were used. This work was extended in [17] where a block-based motion estimation technique was used to find motion vectors between successive images. Such vectors are then rearranged into intensity images and transformed into the frequency domain.

This paper is organized as follows. Section II describes the Arabic sign language database used in the work. Section III describes the hand features. Hand tracking and recognition phases of the proposed system are elaborated in Section IV. Section V presents the modeling of the Arabic Sign Langue using Hidden Markov Models. The experiments and results are discussed in Section VI. Finally, conclusion and future work are presented in Section VIII.

II. ARABIC SIGN LANGUAGE (ArSL) DATABASE

Because there has been no serious attention to Arabic sign language recognition, there are no common databases available for researchers in this field. Therefore, we had to build our own database with reasonable size. As depicted in Fig. 1, in the video capturing stage, a single digital camera was used to acquire the gestures from signers in a video format. At this stage, the video is saved in the AVI format in order to be analyzed in later stages. When it comes to recognition, the

Figure 1.   Sample videos of our Arabic sign language (ArSL) database.

video is streamed directly to the recognition engine.

The database in this work is collected in collaboration with ASDAA' Association for Serving the Hearing Impaired (ASDAA) [18]. The videos are captured from the deaf community who volunteered to perform the signs to generate samples for our study. The database consists of a 20-word lexicon given in Table 1.

No restrictions are imposed on the signer or word length. The words pertain to common situations in which handicapped people might find themselves in. The database itself consists of 45 repetitions of each of the 20 words performed by different signers, 20 of which are used for training and 18 for testing. No restriction is imposed on clothing, background, age or sex of the signer. Moreover, signers are gloves-free and with different signers with different skin colors. It was totally free hands. The deaf sign word is captured using a digital video camera. The frame rate was set to 25 frames per second with a spatial resolution of 640x480.

III. HAND FEATURES

The image features, together with information about their relative orientation, position and scale, are used for defining understated but discriminating view-based object model [19].

We represent the hand by a model consisting of (i) the palm as a coarse scale blob, (ii) the five fingers as ridges at finer

scales and (iii) finger tips as even finer scale blobs as in Fig. 2. We then define different states of the hand model, depending on the number of open fingers.

TABLE I.   RRECOGNITION RATES OF DIFFERENT HMM MODELS WITHDIFFERENT FEATURE VECTOR LENGTHS

Word Number	Arabic Meaning	English Meaning
1	اسكندرية	Alexandria
2	يشرب	Drink
3	يأكل	Eat
4	انا	Me
5	انت	You
6	لغه انجليزية	English
7	جمعية	Association
8	صم وبكم	Deaf
9	لغه عربية	Arabic
10	لغه فرنسية	French
11	كتاب	Book
12	سعيد	Happy
13	ثرثار	Talkative
14	حائر	Confused
15	غيرمستريح	Uncomfortable
16	مدرسة	School
17	مسئول	Responsible
18	مصر	Egypt
19	منزل	Home
20	ينام	Sleep

To model translations, rotations and scaling transformations of the hand, a feature vector is defined to describe different hand features including the global position y), size and orientation and its discrete state.
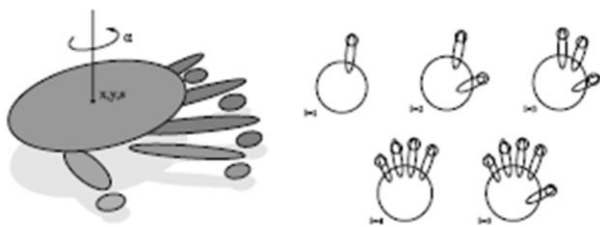

Figure 2. Feature-based hand models in different states. The circles and ellipses correspond to blob and ridge features. When aligning models to images, the features are translated, rotated and scaled according to the feature vector

IV.   HAND TRACKING AND RECOGNITION PHASES OF PROPOSED SYSTEM

In this paper, a system for recognizing Arabic sign language gestures is presented. There are three main phases for hand detection and tracking; skin detection, edge detection and hand fingertips tracking.

A. Skin Detection

Each video contains a collection of frames representing a gesture. At first, each video is pre-processed by applying a video segmentation technique that captures frames with a frame rate of 25Hz. Then, the RGB captured frames are converted into HSV image because it is more related to human color perception [20]. These color spaces separates three components: the hue (H), the saturation (S) and the brightness (V). Essentially, HSV-type color spaces are deformations of the RGB color cube .They can be mapped from the RGB space via a nonlinear transformation. One of advantages of these color spaces in skin detection is that they allow users to specify the boundary of the skin color class in terms of the hue and saturation. As V gives the brightness information, they are often dropped to reduce illumination dependency of skin color.

Given an image, each pixel in the image is classified as a skin or non - skin using color information. The histogram is normalized and if the height of the bin corresponding to H and S values of a pixel exceeds a threshold called skin threshold (obtained empirically), this pixel is considered a skin pixel. Otherwise, the pixel is considered a non-skin pixel. A general image and its skin detected image can be seen such that white pixels represent the hand gesture and black pixels represent the background or any object behind the skin as shown in Fig. 3 (a). Finally, smoothing is applied to each frame using a MEDIAN filter to remove noise and shadow.

B. Canny Edge Detection

The Canny algorithm uses an optimal edge detector based on a set of criteria which include finding the most edges by minimizing the error rate, marke edges as closely as possible to the actual edges to maximize localization and marke edges only once when a single edge exists for minimal response [21].

C. Hand Contours and Fingertips Tracking

Hand tracking is the process of locating a moving hand (or both hands) over time using a camera. For each frame extract, the contours of all the detected skin regions in binary image using connected component analysis are detected. Tests are performed to detect whether the input contour is convex or not. The contour must be simple, i.e. without self-intersections. The signer's head is considered to be the biggest detected region and the moving hand as the second biggest region. Features considered include the position of the head, coordinates of the center of the hand region and direction angle of the hand region. Other features that represent the shape of the hand are also considered and are extracted from changes of image intensities called image motion:

$$I(x, y, t + T) = I\big(x - \varepsilon(x, y, t, T), y(n(x, y, t, T)\big) \qquad (1)$$

Thus, the next frame recorded at time t + 1 can be obtained by moving every point in the current frame, recorded at time t, by suitable amount. The amount of motion $\delta = (\varepsilon, n)$ is called displacement of the point at X = (x, y).

The displacement vector is a function of the image position X , and variations in it are often noticeable even within the small tracking window. We try to find interesting points with big eigenvalues in an image to be added to the feature vector.

These interesting points (corners) are characterized by a large variety in all directions of the vector . By analyzing the eigenvalues of the image pixels, this characterization can be expressed in the following way: we should have two "large" eigenvalues for an interesting point. Based on the magnitudes of the eigenvalues, the following inferences can be made based on this argument:

If $\lambda1\approx0$ and $\lambda2 \approx0$ (the two "interesting points" eigenvalues for an inter, st point) then this pixel (x,y) has no features of interest. In this case we reject the corners with the minimal eigenvalue less than quality Level

If $\lambda1$ and $\lambda2$ have large positive values, then a corner is found. The Shi-Tomasi [22, 23] corner detector directly computes min $(\lambda1, \lambda2)$ because under certain assumptions, the corners are more stable for tracking.

Finally, it ensures that all the corners found are distanced enough on from another by considering the corners (the strongest corners are considered first) and checking that the distance between the newly considered feature and the features considered earlier is larger than the minimum distance. So, the function removes the features than are too close to the stronger feature.

An example of interesting points found that represent a motion is shown in Fig. 3(b).In this work, the implementation of Hand tracking method is carried out using OpenCV (Open Source Computer Vision); a library of programming functions for real time computer vision. [24].

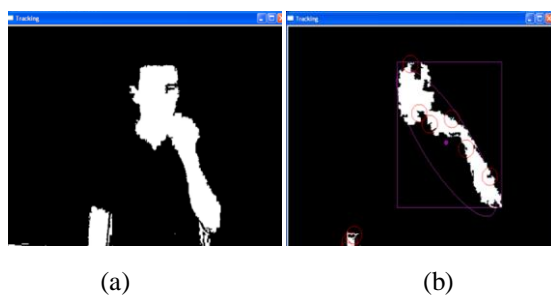(a)   (b)

Figure 5.   the result of computing blob features and ridge features from an image of a hand. (a) Result image after skin detection (b) circles and ellipses corresponding to the significant blob and features extracted from an image of a hand; it describes how the selected image features capture the essential structure of a hand.

V.   MODELLING OF ArSL USING HMM

HMMs (Hidden Markov Models) have been prominently and successfully used in sign languages. HMM is a probabilistic model representing a given process with a set of states (not directly observed) and transition probabilities between the states. Such a model has been used in a number of applications including the recognition of the Sign Language recognition [25, 26].

Let each sign be represented by a sequence of gestures or *observations O*, defined as:

$$O = o1, o2,........., o\tau \qquad (2)$$

Where o_t is the feature vector observed at time t. The sign recognition problem can then be regarded as that of computing:
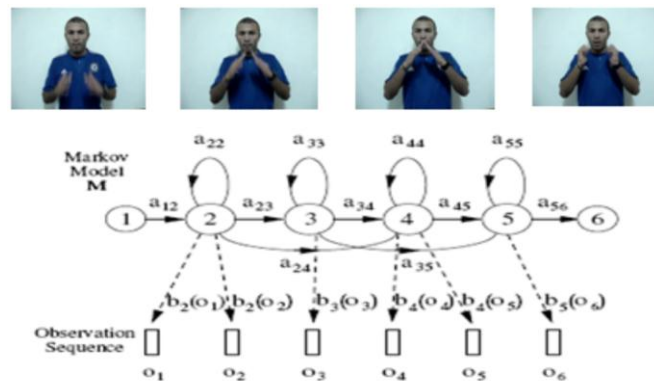
$$arg\ maxi\ \{P(wi|O)\} \qquad (3)$$

Figure 3.   State HMM model for gesture "اسكندرية Alexandria"

Where w_i is the i'th vocabulary word. This probability is not computable directly but using Bayes' Rule [27]:

$$\Pi(\omega_t|O) =(\ \Pi(O|\omega_t)\Pi(\omega_t))/\ \Pi(O)\ \ (4)$$

Thus, for a given set of prior probabilities P (wi), the most likely sign depends only on the likelihood P (O|wi). Given the dimensionality of the observation sequence O, the direct estimation of the joint conditional probability P (o1, o2 …|wi) from examples of sign is not possible. However, if a parametric model of word production such as a Markov model is.

As shown in Fig 4, each gesture for a sign is modeled as a single HMM with N observations per gesture (o1, o2, … ot). In HMM based sign recognition, it is assumed that the sequence of observed feature vectors corresponding to each gesture is generated by a Markov model as shown in Fig 4. A Markov model is a finite state machine which changes state once every time unit and each time t that a state j is entered, a feature vector ot is generated from the probability density bj(ot).

Furthermore, the transition from state i to state j is also probabilistic and is governed by the discrete probability aij . Fig 4 shows an example of this process where the six state model moves through the state sequence X = 1; 2; 2; 3; 4; 4; 5; 6 in order to generate the sequence o1 to o6. It is to be noted that the entry and exit states of a HMM are non-emitting in the Hidden Markov Model Toolkit which is used in this work [28]. This facilitates the construction of composite models as explained in more detail later.

The joint probability that O is generated by the model M moving through the state sequence X is calculated simply as the product of the transition probabilities and the output probabilities. So is for the state sequence X in Fig 4.

A. Training Phase

Training in the context of our work means learning or generating an HMM given a sequence of observations. For each training sequence XT11 , ...,XTn1 , ...,XTN1 of a gesture of class k with N sequences, the image features are prepared and then extracted .

These extracted images are used as feature vectors in the Viterbi [15] training to train a hidden Markov model λk for each gesture as shown in Fig. 5.
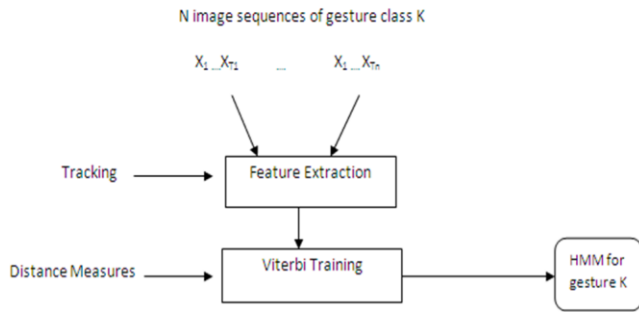

Figure 5.   Traning phase

B. Recognition Phase

This phase involves finding the probability of an observed sequence given an HMM and finding the sequence of hidden states that most probably generated an observed sequence. The feature extraction of the test sequences is identical to the training process. Then for each test pattern the hidden Markov model which best describes the current observation sequence is searched as shown in Fig. 6.
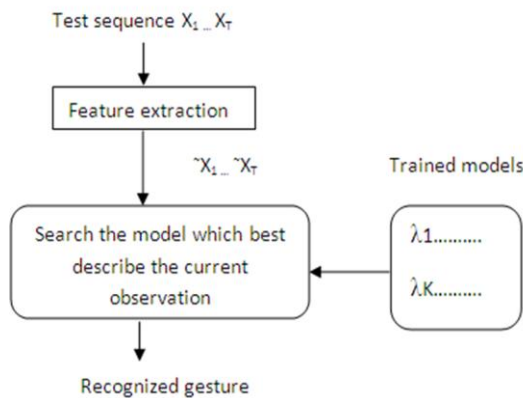

Figure 6.   Recognition Phase

The implementation of the HMM for our ArSL system has been carried out using the HTK toolkit [29]. HTK is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research although it has been used for numerous other research areas including speech synthesis, character recognition and DNA sequencing. HTK is in use at hundreds of sites worldwide.

VI.   EXPERIMENTS AND RESULTS

The training method, described earlier, has been implemented by creating one HMM model per class (gesture), resulting in a total of 20 models.

Several experiments have been conducted to evaluate our ArSL recognition system. All experiments are performed on the same training data collected in prior. Depending on the same database, the system attempts to recognize all samples for every word where the total number of samples considered here is 360. We use 6 HMM models which have different number of states and different number of Gaussian mixtures per state. For each experiment, the 6 models are used for different feature vector lengths (5, 8 and 9) achieving recognition rates of 78.61%, 82.22% and 80.27 % respectively as shown in table 2.

TABLE II.   RECOGNATION RATES OF DIFFERENT HMM MODELS WITH NUMBER OF STATES AND MIXTURE

HMM model	No of features		
	5 elements	*8 elements*	*9 elements*
[3states/10mixture]	47.78%	58.89%	56.67%
[6states/2mixture]	41.94%	45.83%	55.56%
[4states/2mixture]	34.44%	30%	30%
[6states/6mixture]	61.94%	66.11%	68.61%
[6states/4 mixture]	56.94%	56.94%	43.05%
[6states/10mixture]	75.55%	71.94%	74.16%
Overall Recognition Rate	78.61%	82.22%	80.27%

The set of experiments shown in table 2 has been conducted for each of the 20 Arabic signs in our database. The best result (recognition rate) obtained for each sign along with the associated best model is shown in table 3.

It is noticeable that some signs result in particularly low recognition rates. The gesture "eat", for example, has a recognition rate of 55.56% and is mostly classified as "drink". This is due to the fact that the location, movement and orientation of the dominant hand are very similar in both gestures. Therefore, the observation (feature) vectors, o1, o2,......, ot produced from the feature hand tracking phase are most likely very close to each other. Thus, the system will get confused between these two signs and provide relatively higher error rate for these particular gestures. A similar situation occurs with the sign "Me" which has a recognition rate of 66.67% is mostly classified as {You}.

Our ArSL proposed recognition system is based on 8 features per frame which is considered better than the previously published results in the field of ArSL while Jarrah and Halawani [3] use of 30 elements as a length of feature vector per video frame. M. Rousan and K. Assaleh [15] use feature vector of 50 elements.

We compare our work with that done in [28]. In spite of the fact that they use different feature extraction methods, setup, and database, both systems are based on the same classifier (HMM). As shown in Table 4, our proposed system (referred to as ArSL-Using HMM in Table4) system performs much better than the DCT coefficient-based system (referred to as the ArSL – DCT coefficient-based system in Table 4).

ArSL– DCT coefficient-based system uses 50 elements length of feature vector and Recognition rate 90.6%.It is expected that the increase in a feature vector size be accompanied by a corresponding increase in recognition rates. This is due to the fact that each DCT coefficient is uncorrelated with other coefficients and hence no redundant information is presented in increasing coefficients.

TABLE III.   RECOGNITION RATES WITH DIFFERENT MODELS FOR EVERY
WORD WITH BEST MODEL

,Arabic Meaning	English Meaning	Best Result	HMM [Number of States / Number of Gaussian mixtures]
اسكندرية	Alexandria	77.78%	[6states/10 mixture]
يشرب	Drink	88.89%	[3states/10 mixture]
ياكل	Eat	55.56%	[3states/10 mixture]
انا	Me	83.33%	[6states/10 mixture]
انت	You	66.67%	[6states/10 mixture]
لغه انجليزية	English	88.89%	[6states/10 mixture]
جمعية	Association	66.67%	[6states/6 mixture]
صم وبكم	Deaf	88.89%	[6states/10 mixture]
لغه عربية	Arabic	88.89%	[6states/6 mixture]
لغه فرنسية	French	94.44%	[6states/4 mixture]
كتاب	Book	55.56%	[6states/10 mixture]
سعيد	Happy	100%	[6states/6 mixture]
ثرثار	Talkative	72.22%	[6states/10 mixture]
حائر	Confused	100%	[6states/10 mixture]
غيرمستريح	Uncomfortable	88.89%	[6states/6 mixture]
مدرسة	School	77.78%	[6states/10 mixture]
مسئول	Responsible	100%	[6states/10 mixture]
مصر	Egypt	77.78%	[6states/6 mixture]
منزل	Home	72.22%	[6states/6 mixture]
ينام	Sleep	100%	[3states/10 mixture]

TABLE IV.   COMPARISON WITH SIMILAR SIGNER-INDEPENDENT, HMM-
BASED SYSTEMS.

	Instruments used	Feature vector length	Recognition Rate
ArSL–DCT coefficient-based system [28]	None: free hands	50 elements of DCT coefficients per frame	90.6%
ArSL-Using HMM	None: free hands considering the head position	8 features per frame	82.22%

VII. CONCLUSION AND FUTURE WORK

We have demonstrated that our Arabic sign language recognition system is effective considering the nature of the videos used and the number of features considered. Our system is signer-independent. The database that we built consists of videos taken for deaf people using their normal life Arabic sign language. The signers are gloves-free, with varying clothes and skin colors. Importantly, only 8 features have been considered which is less than the number of features used previously by other researchers.   The overall recognition rate is 82.22%, which is reasonably high considering the number of features used.

In the future, we aim to achieve higher recognition rates with a larger data set. A psycholinguistic study on the structure of Arabic sign language might be needed to choose the appropriate HMM model (the perfect number of states) for each gesture. We will explore and test our training models to build a continuous sentence recognition system using a sub-gesture word based recognition system. Such a system will help the deaf community to interact and integrate with the rest of the society.

REFERENCES

[1]   K. Assaleh and M. Al-Rousan, "Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers" EURASIP Journal on Applied Signal Processing, 2005 (13): 2136-2146,2005.

[2]   W. Gao, J.Y. Ma, J.Q. Wu, C.L. Wang, Sign language recognition based on HMM/ ANN/DP, International Journal of Pattern Recognition Artificial Intelligent 14 (5) (2000) 587–602.

[3]   O. Al-Jarrah, A. Halawani, Recognition of gestures in Arabic sign language using neuro-fuzzy systems, Artificial Intelligence 2 (133) 117–138, 2001.

[4]   M.W. Kadous, Machine recognition of Auslan signs using PowerGloves: towards large-lexicon recognition of sign language, in: Proceedings of the Workshop on the Integration of Gestures in Language and Speech, pp. 165–174, 1996.

[5]   K. Grobel, M. Assan, Isolated sign language recognition using hidden Markov models, in: Proceedings of the International Conference on System, Man and Cybernetics, pp. 162–167, 1997.

[6]   H. Hienz, B. Bauer, K.F. Krais, HMM-based continuous sign language recognition using stochastic grammar, in: Proceedings ofGW'99, LNAI 1739, pp. 185–196, 1999.

[7]   B. Bauer, K.F. Kraiss, Towards an automatic sign language recognition system using subunits, in: Proceedings of the International Gesture Workshop, pp.64–75, 2001.

[8]   R.H. Liang, M. Ouhyoung, A real-time continuous gesture recognition system for sign language, in: Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, pp. 558–565, 1998.

[9]   V.I. Pavlovic, R. Sharma, T.S. Huang, Visual interpretation of hand gestures for human–computer interaction: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) 677–695, 1997.

[10] J.J. Triesch, C. Malsburg, A system for person-independent hand posture recognition against complex backgrounds, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (12) 1449–1453, 2001.

[11] P. Vamplew, A. Adams, Recognition of sign language gestures using neural networks,Australian Journal of Intelligent Information Processing Systems 5 (2) 94–102, 1998.

[12] T. Starner, J. Weaver, A. Pentland, Real-time American sign language recognition using desk and wearable computer-based video, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (12) 1371–1375, 1998.

[13] Y. Wu, T.S. Huang, Vision-based gesture recognition: a review, in: Proceedings of the International Gesture Workshop, pp. 103–115, 1999.

[14] G.L. Fang, W. Gao, J.Y. Ma, Signer-independent sign language recognition based on SOFM/HMM, in: Proceedings of the IEEE ICCV Workshop Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, pp. 90–95, 2001.

[15] M. Mohandes, S. I. Quadri, M. Deriche "Arabic Sign Language Recognition an Image–Based Approach" 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07) 0-7695-2847-3/07 $20.00 © 2007.

[16] T. Shanableh, K. Assaleh and M. Al-Rousan, "Spatio- Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language" IEEE Trans. On Systems, Man and Cybernetics Part B, 37 (3): 641-650, 2007.

[17] T. Shanableh and K. Assaleh, "Telescopic Vector Composition and polar accumulated motion residuals for feature extraction in Arabic Sign Language recognition" EURASIP Journal on Image and Video Processing, vol. 2007, Article ID 87929, 10 pages, 2007.

[18] ASDAA' Association For Serving The Hearing Impaired (ASDAA)

[19] Lars Bretzner1, 2 Ivan Laptev1, Tony Lindeberg   "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering" Proceedings of the Fifth IEEE International

Conference on Automatic Face and Gesture Recognition (FGR.02) 0-7695-1602-5/02 $17.00 IEEE © 2002

[20] A. Albiol, L. Torres, and E. J. Delp. "Optimum color spaces for skin detection." In proceedings of the 2001 international conference on image processing, volume 1, vol. 1, pp. 122-124 ,2001.

[21] Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V Pujari "Finger Detection for Sign Language Recognition" Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, Hong Kong, 2009.

[22] J. Shi and C. Tomasi, "Good Features to Track,". 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer, June 1994.

[23] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features". Pattern Recognition 37: 165–168, 2004.

[24] Open Source Computer website. [Online]. Available Vision http://opencv.willowgarage.com/

[25] Prof. Dr.-Ing. H. Ney, "Appearance-Based Gesture Recognition" Diplomarbeit im Fach Informatik Rheinisch-Westf`alische Technische Hochschule Aachen Lehrstuhl f`ur Informatik VI, 2005

[26] Khaled Assaleh, Tamer Shanableh, Mustafa Fanaswala, Harish Bajaj, and Farnaz Amin, "Vision-based system for Continuous Arabic Sign Language Recognition in user dependent mode" Proceeding of the 5th International Symposium on Mechatronics and its Applications (ISMA08), Amman, Jordan, May 27-29, 2008.

[27] Forney GD, "The Viterbi algorithm". Proceedings of the IEEE 61 (3): 268–278. doi:10.1109/PROC.1973.9030, 1973.

[28] M. AL-Rousan, K. Assaleh , A. Tala'a. "Video-based signer-independent Arabic sign language recognition using hidden Markov models" Applied Soft Computing 9 990–999, 2009.

[29] The HTK website. [Online]. Available http://htk.eng.cam.ac.uk

AUTHORS PROFILE

Aliaa A. A.Youssif, professor of computer science. in Faculty of Computers and Information, Helwan University. Cairo, Egypt. She received her B.Sc and MSc. degree in telecommunications and electronics engineering from Helwan University. Dr. A. Youssif received the PhD degree in computer science from Helwan University in 2000. She was a visiting professor at George Washington University (Washington DC, USA) in 2005. She was also a visiting professor at Cardiff University in UK (2008). Her fields of interest include pattern recognition, AI researches, and medical imaging. She published more than 40 papers in different fields.

Amal Elsayed Aboutabl is currently an Assistant Professor at the Computer Science Department, Faculty of Computers and Information, Helwan University, Cairo, Egypt. She received her B.Sc. in Computer Science from the American University in Cairo and both of her M.Sc. and Ph.D. in Computer Science from Cairo University. She worked for IBM and ICL in Egypt for seven years. She was also a Fulbright Scholar at the Department of Computer Science, University of Virginia, USA. Her current research interests include parallel computing, performance evaluation and image processing.

Heba Hamdy Ali received her BSc. degree in computer science from Helwan University. She is currently a master's degree student under supervision of prof. Aliaa A. A.Youssif and Dr Amal Elsayed Aboutabl. Her areas of interests include image processing, pattern recognition, HMM and artificial intelligence.