Supporting Arabic Sign Language Recognition with Facial Expressions

SASLRWFE

Ghada Dahy Fathy
Faculty of Computers and Information
Cairo University
Cairo, Egypt
g.dahy@fci-cu.edu.eg

E.Emary
Faculty of Computers and Information
Cairo University
Cairo, Egypt
e.emary@fci-cu.edu.eg

Hesham N. El Mahdy
Faculty of Computers and Information
Cairo University
Cairo, Egypt
ehesham@cu.edu.eg

Abstract— this paper presents an automatic translation model for the combination of facial expressions of user and gestures of manual alphabets in the Arabic sign language. The part of facial expressions depends on locations of user's mouth, nose and eyes . The part of gestures of manual alphabets in the Arabic sign language does not rely on using any gloves or visual markings to accomplish the recognition job. As an alternative, it deals with images of signer's hands. Two parts enable the user to interact with the environment in a natural way. First part in the model deals with signs and consists of three phases preprocessing phase, skin detection phase and feature extraction phase. Second part in the model that deals with facial expressions consists of two phases face detection and tracking facial expression. Proposed model has an accuracy 90% using minimum distance classifier (MDC) and absolute difference classifier in case of facial expressions and 99% in case of signer's hands**.**

Keywords— ***Arabic Sign Language, Facial Expression, Minimum Distance Classifier (MDC), Human computer interaction (HCI), Absolute Distance Classifier (ADC).***

I. INTRODUCTION

Signing has always been part of human communications. Sign language (SL) is a form of manual communication and is one of the important communications for people in deaf community (1). For thousands of years, deaf people have generated and used signs among themselves. In the past, signs were the only ways of communication available for all deaf people. The sign language is the fundamental communication method between people who suffer from hearing impairments. As we know about oral language, sign language is not universal because it has different features that differ from country to other according to the country, or even according to the regions. Sign language in the Arab World has recently been recognized and documented. Very great efforts have been made to build the sign language used in individual countries, including Jordan, Egypt and the Gulf States, by trying to standardize the language and spread it among members of the deaf environment. In the recent years, the idea of the computerized translator became an interesting research area (2). There are two ways for interacting between human and computer: glove-based and vision-based systems (3). The glove-based system depends on electromechanical devices that are used for data collection about the gestures (4). The user has to wear some sort of gloves that cover with sensors to make the interaction between the system and computer. According to readings of sensors signs meaning will be understood. There is difficult for signers in moving with great numbers of sensors so second way of human computer interaction (HCI) has been provided to overcome this problem. Second way depending on image of signers in their communication use two channels: manual and non manual. In the manual channel, deaf people use their hands to express lexical meaning. In the non manual channel deaf people use their facial expression, upper body movements and head to express syntactic and semantic information. Non-manual expression co-occurs with manual signs to support users. In this work, our goal is to construct a model that is able to translate Arabic sign language (ASL) to Arabic text. We take in our consideration grammatical expressions that provide the grammatical structure of sentence. We use four face emotions for dealing with non-manual expression neutral, sad, happy, and angry. Each type composed of a combination of facial features movements. For identifying facial expression in sign language we tracked sets of features in faces image like eyes, nose and mouth locations. The paper is composed of six main sections. First section will be about related works that discuss previous work in sign language and facial expressions second section will be about proposed schema model and how we extract features of signs and facial expressions. Third section will be about methodology in sign language part. The fourth section will be about methodology in facial expressions part. Experimental results will be discussed in section number five. The last section will contain summary about paper and future work.

II. RELATED WORK

In recent years, several research projects in developing sign language systems were presented (5). An Arabic Sign Language Translation Systems (ArSL-TS) (6) model has been introduced. This model for sign language runs on mobile devices that model enable users to translate Arabic text into Arabic Sign Language for the deaf on mobile devices such as Personal Digital Assistants (PDAs). Software in (7) consists of two basic modules: linguistic translation from printed English into sign language, and virtual human animation. The animation software enables Simon to sign in real-time. A dictionary of signed words makes system to look up the accompanying physical movement, facial expressions and body positions, which are stored as motion-capture date on a hard disk. This model contains very realistic and accurate hand representations, developed within the project. Moreover, natural skin textures are applied to the hands and face of the model to generate the maximum impression of subjective reality. In (8), an automatic Thai finger-spelling sign language translation system was developed using Fuzzy C-Means (FCM) and Scale Invariant Feature Transform (SIFT) algorithms. Key frames took from several subjects at different times of day and for several days. Also, testing Thai finger spelling words video took from 4 subjects with the SIFT threshold of 0.7 and use one nearest neighbor prototype. In (9), an automatic translation of static gestures of alphabets in American Sign Language (ASL) was developed, ASL used three feature extraction methods and used neural network to classify signs. The proposed system interacts with images of bare hands, which allows user to interact with environment in as normal people. Token image would be processed and converted to a feature vector that will be compared with the feature vectors of a training set of signs. The system is implemented and tested using data sets of hand images samples for each signs. System used three feature extraction methods are tested and the best method is suggested with results obtained from Artificial Neural Network (ANN). Recent works on tracking facial features used sets of Active. Shape Models to constrain face shapes and also considered head motions (9), (10). KLT was used in (11) to track facial feature points, but it had problem because their 2D local models for shape constraints that were based on frontal face might not cope well under varying head pose.

Algorithm 1 Facial Feature Extraction.

1: Get frames that contain facial movement.
2: Apply median filter with 3×3 windows to remove noise from frames.
3: Convert RGB image into YCbCr to detect skin.
4: Calculate first component in YCbCr

$$Y = 16 + (65.481.R + 128.553.G + 24.966.B). \quad (1)$$

5: Calculate second component in YCbCr

$$C\,b = 128 + (-37.797.R - 74.203.G + 112.0B). \quad (2)$$

6: Calculate third component in YCbCr

$$C\,r = 128 + (112.0.R - 93.786.G + 18.214.B). \quad (3)$$

7: Mark skin pixel to detect face that contain $Cb \geq 77, Cb \geq 127, Cr \geq 133$ and $Cr \geq 173$.
8: Detect boundaries by using Sobel (15) after that applying horizontal projection to mark eyes region. Taking the upper half of face and calculate the vertical projection to separate eyebrows from eyes.
9: Select the lower part of face and calculate vertical projection to get mouth and nose region.
10: Draw rectangular box on each of the detected feature Elements.
11: Generate feature vector of width and height of each rectangular.

Algorithm 2 Signs Feature Extraction

1: Get video that represent hand movement.
2: Divide video into frames.
3: Apply median filter with 3×3 windows to remove noise from frames.
4: Convert RGB image into YCbCr to detect skin.
5: Calculate first component in YCbCr

$$Y = 16 + (65.481.R + 128.553.G + 24.966.B) \quad (4)$$

6: Calculate second component in YCbCr

$$C\,b = 128 + (-37.797.R - 74.203.G + 112.0B) \quad (5)$$

7: Calculate third component in YCbCr

$$C\,r = 128 + (112.0.R - 93.786.G + 18.214.B) \quad (6)$$

8: Mark skin pixel that contain Cb≥77, Cb≥127, Cr≥133 and Cr≥173 and crop image that contain skin.
9: Divide cropped image into blocks each block with size 4×4 pixels.
10: Get centroid of block

$$\text{Centroid of block} = \frac{\sum_{i=1}^{16} \text{Element of block number i}}{16} \quad (7)$$

11: Store all centroids of all blocks as extracted features

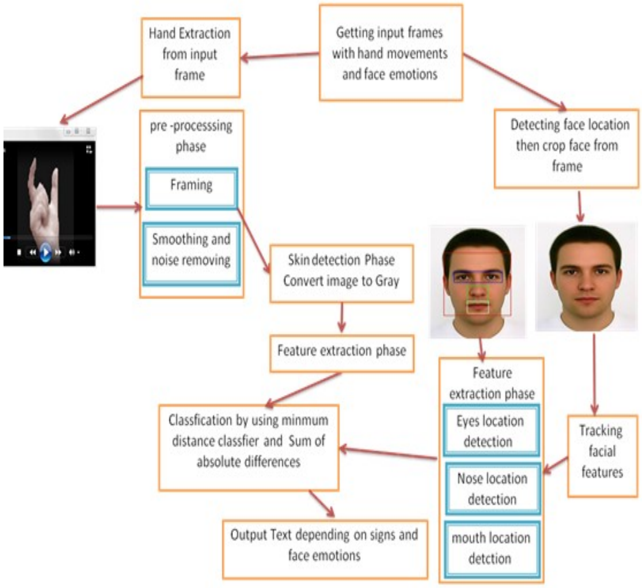

Fig. 1: Facial expressions and Arabic sign language model architecture.

III. THE PROPOSED APPROACH

Facial expressions in sign language model are composed of three main phases for feature extraction of signs namely, Pre- processing phase, Skin detection phase and Feature extraction. Model composed of 2 main phases for facial expressions extraction namely, face detection and tracking facial features. Figure 1 depicts the structure of the Facial expressions in sign language model. Pre-processing phase in the signs part receives, as an input, a video that contains the signed words to be translated into text and prepare it to be ready for use in subsequent phases. Skin detection phase in the signs part detect skin in image by converting RGB image into YCbCr formatting. YCbCr is a family of color spaces. YCbCr has better accuracy compared with other color spaces families in case of skin detection. YCbCr presents color as bright- ness and two color difference signals. Components Y is the brightness (luma), Cb and Cr are two colors Difference signals. Model calculates YCbCr components by using equations in algorithm 2 after that model converts input frame into gray to enables us in defining and separating location of hands and background. Finally we extract features from input frame. Detecting and cropping face in the second part of the system is very important phase. For face detection we use YCbCr color space model to define the location of face in the image. Values of CB and Cr component support model in defining the skin part in the input frames as we shown in algorithm 1. In classification phase, each unknown facial expressions or signs are being matched with all the known expressions and signs in the same category in the database and takes the nearest one to expressions. Database of the model deals with 7 facial expressions neutral, smile, sad, angry, afraid, disgusted and surprised. There are 105 samples of facial expressions. 15 samples for neutral face, 15 samples for smile face, 15 samples for sad, 15 sample for angry face, 15 samples for afraid face, 15 samples for disgusted, 15 samples for surprised face. Database of the model contains dictionary for all Arabic signs.

IV. METHODOLOGY: PHASE-I

A. Pre-Processing

Firstly, a video that contains stream of signed words (gestures) to be translated is acquired. After that, the video enters the pre-processing phase where video divides into frames. Then, smoothing is applied for each frame to remove noise by using median filter with 3X 3 windows. The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value(If the neighborhood under consideration contains an even number of pixels, the average of the two middle pixel values is used).

B. Skin Detection

In that phase system tries to detect the skin part in the input frame because the skin part represents hands in the frame

Firstly system converts RGB image into YCbCr image. System calculates YCbCr components by using equations in algorithm 2. In figure 2, we see the difference between original image and YCbCr image. Finally system converts the frame into gray by using equation number 8 to isolates skin with black color from background as shown in figure 3.

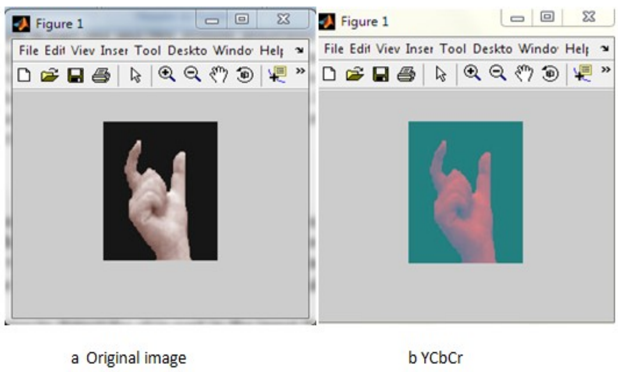

Fig. 2: Converting Original image into YCbCr.

$$Gray(x) = 0.29R + 0.59G + 0.11B. \quad (8)$$

- Where x is the input pixel.
- R is the red value of input pixel.
- G is the green value of input pixel.
- B is the blue value of input pixel.

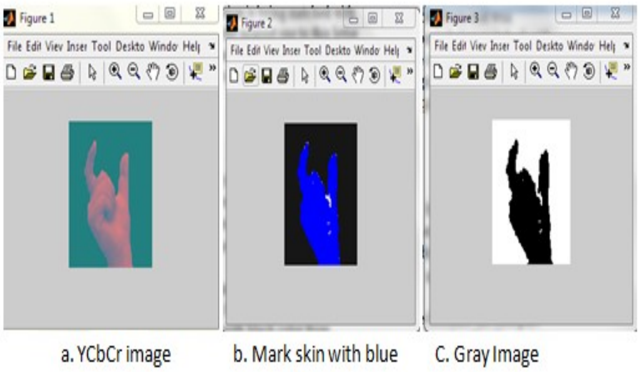

Fig. 3: Skin detection process

C. Feature Extraction

Feature extraction phase depends on Centroid. Firstly sys- tem divides the input frame into blocks with size 4X 4. In that model we use centroid properties for extracting features from blocks as we shown in equation 7 in algorithm 2.

V. METHODOLOGY: PHASE-II

A. Detecting and cropping phase

Recognition algorithms divide into two main approaches, geometric, that depends on distinguishing features, or photometric, which is a statistical approach that distills an image into values and compares the values with templates to eliminate variances. Popular recognition algorithms include Principal Component Analysis using eigenfaces, Linear Discriminate Analysis, Elastic Bunch Graph Matching using the Fisher face algorithm, the Hidden Markov model, the Multilinear Subspace Learning using tensor representation, and the neuronal motivated dynamic link matching. System use YCbCr color space model to define the location of face in the image. Values of CB and Cr component support system in defining the skin part in the input frames as we shown in algorithm 1. We use the skin part in defining face location and drawing rectangle around it as shown in figure 4.

B. Tracking facial features

After detecting face location system use manual way to track facial features. System able to detect eyes, nose and mouth By using vertical projection in the upper and lower part of detected face as shown in algorithm 1 and figure4.

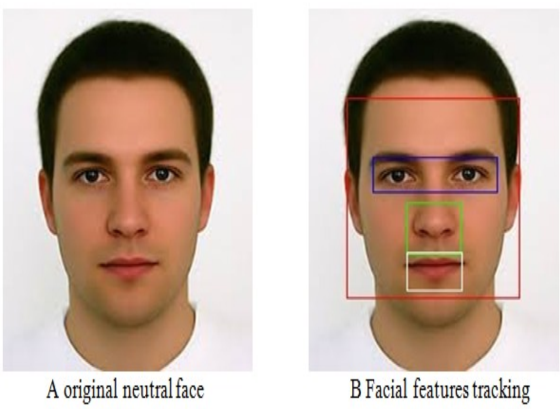

Fig. 4: Extracting facial features

VI. CLASSIFICATION

System stores dictionary for extracted features of Arabic sign language in database. That dictionary supports it in classification. Also in database contains extracted features of face emotions. We use in classification two classifier minimum distance classifier and sum of absolute difference classifier.

A. Minimum Distance Classifier

The minimum distance classifier (MDC) is an example of known used conventional classifier (12), (13). The single nearest neighbor technique completely bypass the problem of probability distance and simply classifies any unknown sample as belonging to the same class of the most similar or nearest Feature vector in the training set of data. Nearest will be taken to the smallest Euclidean distance in dimensional feature space and the classifier compares the extracted new feature vector x(i) with all the class known feature vectors y(i) and choose the class that minimizes the distance classifier using equation 9

$$\text{Distance} = \sum_{i=1}^{N} |(y(i) - x(i))| \qquad (9)$$

Where N is the feature vector length

B. Sum of absolute difference classifier

Sum of absolute difference classifier is considered as a single nearest. It depends on absolute distance between the new feature vector x(i) with all the class known feature vectors y(i) using equation number 10.

$$\text{Distance} = \sum_{I=1}^{N} (y(i) - x(i))2 \qquad (10)$$

Where N is the feature vector length

VII. EXPERIMENTAL RESULTS

In the first part in the model, we use Arabic dictionary for all alpha characters as shown in figure 5. To evaluate the performance of the first part, several videos containing sequences of letters such as "Noon, Ayn, Miem" to generate "Nam" word and "la" have been classified. The system detected the "Noon, Ayn, Miem,la" and generate ”Nam” word and ”la” word. It has accuracy 99% as we shown in table I. It has the best accuracy comparing with other systems accuracy as we shown in figure 6.

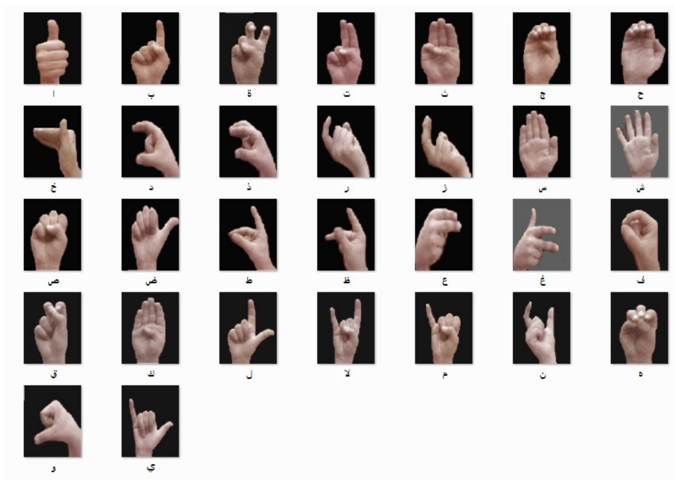

Fig. 5: Arabic signs alpha dictionary

TABLE I: Arabic sign language recognition

Sign language	Classifier	Recognition rate
Arabic Sign Language	MDC	91.3%
Arabic Sign Language	multilayer perceptron	83.7%
Video-based (3)	hidden Markov	93.8%
Our paper	MDC	99%
Our paper	ADC	99%

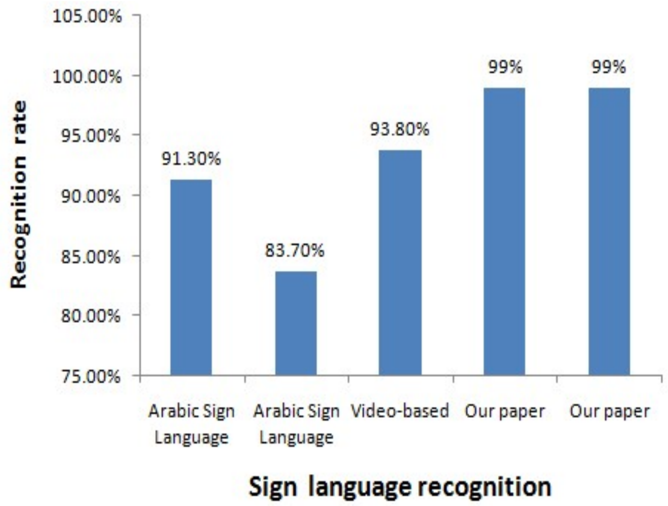

Fig. 6: Arabic sign language recognition rate

In the second part of the system, we are trying to detect face in the image. For face detection we use OpenCv library to support us in defining face location. OpenCV is released under a BSD license; OpenCV is free for both academic and commercial using. It has C++, C, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android. OpenCV was designed for computational efficiency and with a strong support for real-time applications. OpenCv. Implementation in C++/C library can take advantage of multi- core processing. Enabled with OpenCL, it takes advantage of the hardware acceleration of the underlying heterogeneous compute platform. Adopted all around the world or a video frame from a video source. One of its common ways to do this is comparing selected facial features from the image and a facial database. Some of popular facial recognition algorithms identify facial features by extracting landmarks, or features, from an image of the face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw these features are then used to search for other images with matching features. There are other algorithms that depend on normalizing a gallery of face images and compress the face data, save only the data in the image that is important for face recognition. A probe image is then compared with the face data. One of the most successful systems is depended on template matching techniques applied to a set of salient facial features providing a sort of compressed face representation. In that system we use Opencv library which contains haarcas- cade frontalface objects that depending on popular algorithm for defining face location. OpenCV support us in defining eyes, mouse and nose location. Depending on defined locations system able to generate feature vector of width and height of each feature location. System deals with seven facial expressions neutral, smile, sad, angry, afraid, disgusted and surprised. We take 105 straight samples from The Karolinska Directed Emotional Faces (KDEF) for training and 30 straight samples for testing. KDEF is a set of totally 4900 pictures of human facial expressions of emotion. The material was developed in 1998 by Daniel Lundqvist, Anders Flykt and Professor Arn at Karolinska

Institutet, Department of Clinical Neuroscience, Section of Psychology, Stockholm, Sweden.we took 15 training samples for each emotion as we shown in figure 7 the training samples of smile facial expression.

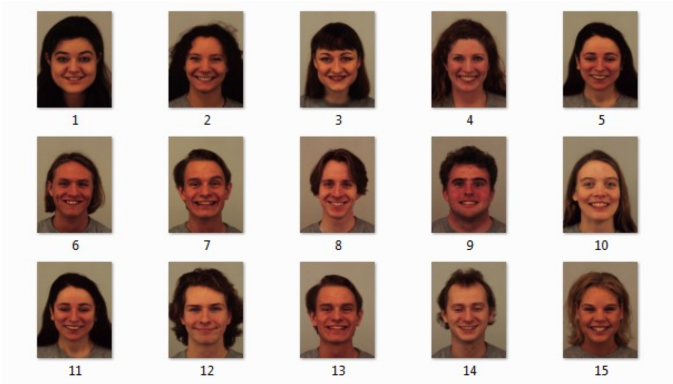

Fig. 7: Training samples of smile face emotion

In case of testing, we use 30 samples from KDEF database for testing. Result of testing by using minimum distance classifier was 90% and also 90% by using absolute difference classifier as we shown in table II and table III. System has the best accuracy comparing with other systems as we shown in figure 8 and table IV.

TABLE II: Error Matrix

	Neutral	Smile	Sad	Angry	Afraid	Disgusted	Surprise
Neutral	4	0	0	0	0	0	1
Smile	0	4	0	0	0	0	0
Sad	0	0	5	0	0	0	0
Angry	0	0	0	4	1	0	0
Afraid	0	0	0	1	1	0	0
Disgust	0	0	0	1	1	3	1
Surprise	1	0	0	0	0	1	6

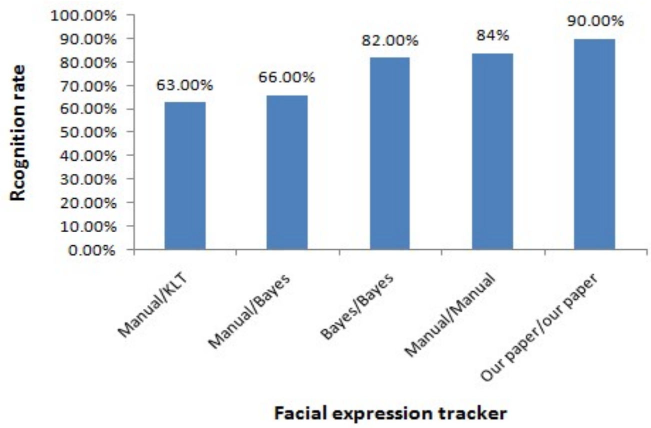

Fig. 8: Facial expressions tracker.

TABLE III: Testing input facial expression

Emotion	No of tested samples	Results	Error
Neutral	5	4	one
Smile	4	4	zero
Sad	5	5	zero
Angry	4	5	one
Afraid	2	1	one
Disgusted	4	3	one
Surprised	6	8	one

TABLE IV: Recognition rate that result from using different facial expression tracker in training and testing.

Training Tracker	Testing tracker	Number	Recognition rate
KLT tracker (14)	KLT tracker	4	76%
Manual tracker (14)	KLT tracker	4	63%
Manual tracker (14)	Bayes tracker	4	66%
Bayes tracker (14)	Bayes tracker	4	82%
Manual tracker (14)	Manual tracker	4	84%
Our paper tracker	Our paper tracker	7	90%

TABLE V: Final decision depending on the final result from first subsystem of signs and second subsystem of facial expressions

Emotion	Signs	Final decision	Signs	Final decision
Neutral	Nam	Nam	La	La
Smile	Nam	Nam	La	Nam
Sad	Nam	La	La	La
Angry	Nam	La	La	La

VIII. CONLUSION AND FUTURE WORKS

In this paper, a system for the purpose of the recognition and translation of the alphabets in the Arabic sign language was implemented. System takes facial expressions in its considerations in translation. System consists of two parts, first part for manual signs and consists of three phases preprocessing phase, skin detection phase and feature extraction phase. Second part in the system that deals with facial expressions consist of two phases detects face and tracking facial expression. System has an accuracy of 90% using minimum distance classifier (MDC) and absolute difference classifier in facial expressions extraction and 99% in case of signs extraction. In the future we will add additional improvements to system to be used for mobile applications to provide easy communication way among deaf/hearing-impaired people. We also could be developed to be provided as a web service used in the field of conferences and meetings attended by deaf people. That system can be used in intelligent classrooms and intelligent environments for real-time translation for sign language. We can support system with other facial expressions like afraid, disgusted and surprised. Common grammatical expressions like Yes/no question (YN), Wh question (WH), Topic (TP), and Negation (NEG) can be developed and add to system to save time and add more supporting to deaf people in their communication. We will increase the size of the database for training and testing. We will use different direction of faces to represent facial expression in 3D. We will use others classifier in testing.

REFERENCES

[1] Nashwa El-Bendary1, Hossam M. Zawbaa2, Mahmoud S. Daoud2,Aboul Ella Hassanien2, and KazumiNakamatsu3, ”Arabic Sign Language Alphabets Translator”,International Journal of Computer Information Systems and Industrial Management Applications. Vol. 3, No. 2, PP. 498-506, 2011.

[2] O. Al-Jarrah and A. Halawani.”Recognition of Gestures in Arabic Sign Language Using Neuro-Fuzzy Systems”, Artificial Intelligence, Vol. 133, No. 1-2, PP. 117-138, 2001.

[3] M. AL-Rousan, K. Assaleh, and A. Tala.”Video-based Signer-independent Arabic Sign Language Recognition Using Hidden Markov Models”, Applied Soft Computing, Vol. 9, No. 3, PP. 990-999, 2009

[4] V. I. Pavlovic, R. Sharma, and T. S. Huang. ”Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review”, IEEE Trans. Pattern Anal. Machine Intell. , Vol. 19, No. 7, PP. 677-695, 1997

[5] M. Huenerfauth. ”Generating American Sign Language Classifier Predicates For English-To ASL Machine Translation”, Ph.D dissertation, University of Pennsylvania, Department of Computer and Information Science, Philadelphia, PA, USA, 2006.

[6] S. M. Halawani. ”Arabic Sign Language Translation System on Mobile Devices”, International Journal of Computer Science and Network Security (IJCSNS), Vol. 8, No. 1, PP. 251-256, 2008

[7] J.A. Bangham, S.J. Cox, M. Lincoln, ITutt, and M. Wells.”Signing for the Deaf Using Virtual Humans”, IEE Seminar on Speech and Language Processing for Disabled and Elderly People, No. 2000/025, PP. 4/1-4/5, 2000.

[8] S.Phitakwinai, S. Auephanwiriyakul, and N.Theera-Umpon.”Thai Sign Language Translation Using Fuzzy C-Means and Scale Invariant Feature Transform”, The Proceedings of International Conference of Computational Science and Its Applications, PP. 1107-1119,Thai,June 2008.

[9] A. Kanaujia and D. N. Metaxas. ”Large Scale Learning of Active Shape Models”.The Proceedings of IEEE International Conference on Image Processing, PP. 265-268,San Antonio,16-19 Sept. 2007.

[10]Y. Tong, Y. Wang, Z. Zhu, and Q. Ji. Robust ”facial feature tracking under varying face pose and facial expression” .Pattern Recognition, Vol. 40, No. 40, PP. 3195-3208, 2007.

[11] Y. Tian, T. Kanade, and J. Cohn.”Recognizing Action Units for Facial Expression Analysis”. IEEE transactions on Pattern Analysis and Machine Intelligent, Vol. 23, No. 2, PP 97-115, 2001.

[12] M. S. Packianather and P. R. Drake. ”Comparison of Neural and Minimum Distance Classifiers in Wood Veneer Defect Identification"', The Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Sage Publications, Vol. 219, No. 11, PP. 831, 2005.

[13] R. Boveiri. ”Persian Printed Numerals Classification Using Extended Moment Invariants”, The proceedings of WASET Int. Conf. on Image and Vision Computing de Janeiro, PP. 167-74,World Academy of Science, Engineering and Technology, 2010.

[14] R. Boveiri. Tan Dat Nguyen and Surendra Ranganath,”Tracking facial features under occlusions and recognizing facial expressions in sign language”,The proceeding of IEEE International Conference,PP. 1 - 7,Amsterdam,17-19 Sept. 2008.

[15] Mr. Manoj K.Vairalkar and Prof. S.U.Nimbhorka. ”Edge Detection of Images Using Sobel Operator”, International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 1, PP. 2250-2459, 2012