Hand Gesture Recognition for Sign Language: A New Higher Order Fuzzy HMM Approach

Saad M. Darwish, Magda M. Madbouly, and Murad B. Khorsheed

Abstract—Sign Languages (SL) are the most accomplished forms of gestural communication. Therefore, their automatic analysis is a real challenge, which is interestingly implied to their lexical and syntactic organization levels. Hidden Markov models (HMM's) have been used prominently and successfully in speech recognition and, more recently, in handwriting recognition. Consequently, they seem ideal for visual recognition of complex, structured hand gestures such as are found in sign language. In this paper several results concerning static hand gesture recognition using an algorithm based on Type-2 Fuzzy HMM (T2FHMM) are presented. The features used as observables in the training as well as in the recognition phases are based on Singular Value Decomposition (SVD) that optimally exposes the geometric structure of a matrix. SVD is an extension of Eigen decomposition to suit non-square matrices to reduce multi-attribute hand gesture data to feature vectors. In our approach, we replace the basic HMM arithmetic operators by some adequate Type-2 fuzzy operators that permits us to relax the additively constraint of probability measures. Therefore, T2FHMMs are able to handle both random and fuzzy uncertainties existing universally in the sequential data. Experimental results show that T2FHMMs can effectively handle noise and dialect uncertainties in hand signals besides a better classification performance than the classical HMMs. The recognition rate of the proposed system is 100% for uniform hand images and 95.5% for cluttered hand images.

Index Terms—Hand gesture recognition, sign language, type-2 fuzzy logic, hidden markov model.

I. INTRODUCTION

A gesture is defined as an energetic movement of hands and creating signs with them such as alphabets, numbers, words and sentences. Gestures are classified into two type static gestures and dynamic gestures. Static gestures refer to certain pattern of hand and finger orientation whereas dynamic gestures involve different movement and orientation of hands and face expressions largely used to recognize continuous stream of sentences [1].

The aim of the gesture recognition is to enable humans to interact with the human made machines in natural way without any mechanical devices and the mathematical equations will be the translator that translates the poses between the gestures and the telerobotic [1]. The gesture recognition is very difficult and complex task since the full

Saad M. Darwish, Magda M. Madbouly are with the Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, 163 Horreya Avenue, El-Shatby 21526, P.O. Box 832, Alexandria, Egypt (e-mail: Saad.darwish@alex-igsr.edu.eg, mmadbouly2@gmail.com).

Murad B. Khorsheed is with the Department of Computer, College of Economic and Administration, University of Kirkuk, Kirkuk, Iraq (e-mail: muradbahram@yahoo.com).

recognition system should be able to identify the hand in different scales, positions, orientations, contrasts, luminosity, and others. Automatic sign languages recognition can be of great significance for communication with deaf people and has also its application in virtual reality, machine control in the industrial field [2].

In general, we can totalize the basic operations for gesture recognition system into four main operations; collecting the inputs, image preprocessing, feature vector extracting and efficient classification algorithm [1]. In the first stage, the meaningful gestures that will be used as an input for the gesture recognition system must be predisposed and decided. Segmentation and tracking are essential in order to extract useful information from raw gesture images. Thus, it is necessary to be able to recognize the region of foreground and split it from the background in a given gesture image. In the third stage, the features (geometric and non-geometric) must be extracted; these features will be used at the time of testing operation. In the final stage, the main job of the classifier is to identify which trained class the current presented testing gesture belongs to. Also the most important thing in hand gesture recognition system is the input features and the selection of good features representation [3].
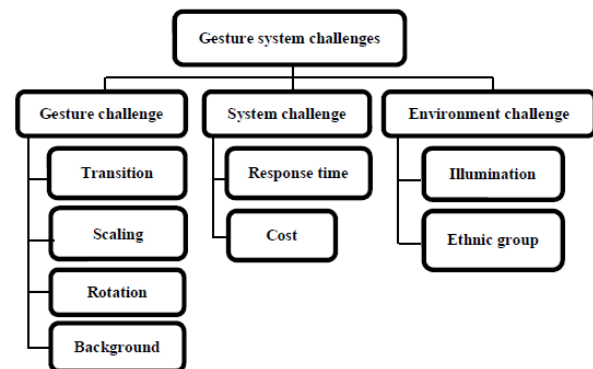

Fig. 1. Gesture challenges.

There are some challenges that are facing the process of gesture recognition and these challenges can be found in preprocessing step and these are common for almost all the gesture recognition application [1], and if the gesture is not well obtained then this will affect all the latter pending process. The challenges can be summarized by the Fig. 1 the gesture recognition is very difficult and complex task since the full recognition system should be able to identify the hand in different situations [3]. To further increase the recognition rate and make the recognition system resilient to view-point variations, the concept of derived features (prominent features are derived from the available data set of features using forward selection algorithm) is introduced

[4]. Readers looking for more information can refer to [1].

Gesture recognition schemes can be broadly classified into two groups. In the first group, a gesture is modeled as a time sequence of states. Here, one uses Hidden Markov models (HMM), discrete finite state machines (DFA), and variants thereof for gesture recognition [5]. In the second group, one uses dynamic time warping to compensate for the speed variations (undulations in the temporal domain) that occur during gesticulation. Gesture recognition schemes can also be categorized on the basis of the parameters that are used to model the appearance of the hand e.g., hand silhouette-based model, graph-based model, use of Fourier descriptors, b-splines, etc. [5]. HMM-based framework is used for robust estimation the individual recognition units from the feature sequence; so that a sequence of recognition units is interpreted as a meaningful gesture. However, exact delimiting of recognition units is essential for good results [6].

In general, vision based hand gesture interface has been attracting more attentions due to no extra hardware requirement except camera, which is very suitable for ubiquitous computing and emerging applications. Methods for vision based hand gesture recognition fall into two categories [6]: 3D model based and appearance model methods. 3D model may exactly describe hand movement and its shape, but most of them are computational expensive to use. In this paper, we focus on appearance model method that uses image features to model the visual appearance of the hand and compare these parameters with the image features extracted from the input images. In general, the performance of vision based gesture interaction is prone to be influenced by illumination changes, complicated backgrounds, camera movement and specific user variance.

Many researchers have made effective efforts to deal with above problems. In this paper, inspired by the work of Jia Zeng and Zhi-Qiang Liu [7] we present a robust gesture recognition engine that identifies a gesture based upon the sequence of hand features. The recognition process involves extracting of the gesturer's hand features through Singular Value Decomposition (SVD), which is very powerful and useful matrix decomposition, particularly in the context of data analysis and dimension reducing transformations [8]. The hand region is recognized by type-2 fuzzy HMM that uses shape information for recognition of the gesture. T2FHMM can effectively handle both randomness and fuzziness of the extracted features for the hand. The system will be designed to maximize the recognition ratio for gesture database under unconstrained environments. We need to consider the trade-off between the computation complexity and robustness.

The rest of the paper is organized as follows. Section II describes some of the recent related works. Section III describes the proposed system. The test results and discussion of the meaning are shown in Section IV. A short summary of this paper and outlook of future work are given in Section V.

II. RELATED WORK

In this section we focus our discussion of the efforts made by researchers on sign language recognition in general. Sign language recognition systems can be further classified into signer-dependent and signer-independent. Accordingly sign language recognition systems are classified into two broad categories: sensor glove based, and vision based systems [9]. The first category requires signers to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation during processing. Glove based methods suffer from drawbacks such as the signer has to wear the sensor hardware along with the glove during the operation of the system. In comparison, vision based systems use image processing algorithms to detect and track hand signs as well as facial expressions of the signer, which is easier to the signer without wearing gloves. However, there are accuracy problems related to image processing algorithms which are a dynamic research area.

There is an extensive literature about methods and systems for static sign language recognition in particular, such as, e.g.: systems for the recognition of 3D and 2-D gestures captured by different devices (data gloves, cameras etc.), methods based on fuzzy logic and fuzzy sets, neural networks, hybrid Neuro-fuzzy methods, etc. [10]. These algorithms perform well in a controlled lighting condition, but fails in case of illumination changes, scaling and rotation. To overcome the problem of multi-variations some popular techniques like Haar-like features, Adaboost classifiers, active learning and appearance based approaches are used [4]. However, all these algorithms suffer from the problem of computational complexity. To increase the accuracy of the hand gesture recognition system, combined feature selection approach is adopted.

During the last decade the methods with potential applications in gestural interfaces for human computer interaction (HCI) are based on HMM that is fundamental for most gesture recognition algorithms used at present. For example, the research in [11] was focused on the application of the HMM method to hand gesture recognition. Since hand images are two-dimensional, it is natural to believe that the 2-DHMM, an extension to the standard HMM, will be helpful and offer a great potential for analyzing and recognizing gesture patterns. However a fully connected 2-DHMMs lead to an algorithm of exponential complexity. The suggested P2-DHMMs use observation vectors that are composed of two-dimensional Discrete Cosine Transform (2-D DCT) coefficients. In addition, their gesture recognition system uses both the temporal and characteristics of the gesture for recognition. Unlike most other schemes, their system is robust to background clutter, does not use special glove to be worn and yet runs in real time.

Another group of researches in this area are based on skin color cues to extract the shape and orientation of the hand. This information is combined with the hand motion estimates, and analyzed using a bank of HMMs to recognize the gesture performed [12]. The system, however, is not very robust to background clutter, and structured noise. While other authors described hand shape using normalized Fourier descriptors. A radial basis function network is used to map the observed hand shape to a set of five predefined shapes. This shape information along with motion information (of the centroid of the binary hand image) is given to an HMM bank to estimate the gesture.

An American Sign Language (ASL) recognition system

developed using a multi-dimensional HMM based method is described in [13]. The authors used a multi-dimensional HMM for better recognition rates compared with one dimensional HMM. In this model, each dimension of the HMM state corresponds to the data from each sensor channel. The multiple data streams from the sensory glove and 3-D motion tracker are the inputs to the HMM process, and this raw data corresponds to the sequence of observations in the HMM. The system can perform online training and real-time recognition of ASL alphabets and basic hand shapes. The evaluation results showed that their proposed method allows fast training and online learning of new gestures and reliable recognition of the trained gestures afterwards.

In [2], a fuzzy classifier has been used to recognize Pakistani Sign Language (PSL). Marked color glove has been used to segment hand and then these marks are also used to extract features to be used by the classifier to recognize sign. The only feature has been used by the fuzzy inference system (FIS) is angle between the fingertip and finger-joint. These angles have the capability to uniquely represent each sign. Results have shown that this technique is capable of recognizing a large dictionary of signs without any prior training and with very low computation requirements.

Based on the publisher papers, there exist some important issues in the design of automatic hand gesture recognition system which if suitably addressed, lead to the development of more efficient recognizers. One of these issues is the extraction of the features. In this paper we investigate the performance of T2FHMM model in hand gesture recognition from color images containing gestures in order to break through the limitations of traditional HMM implementations. The scheme is aimed at improving the performance of the gesture recognizer by presenting a smaller input feature vector through using SVD as a method of generating the HMM observable vectors.

III. Proposed T2FHMM Recognition System

Our method of gesture recognition is a vision based technique that does not use motion sensor gloves or colored gloves for the system to recognize hand shapes. The recognition process involves detecting of the gesturer's hand, Then the detected hand features is determined by SVD analysis. After that T2FHMMs based approach uses shape information that are embedded into SVD for recognition of the gesture. Fig. 2 shows the block diagram of the proposed IT2FHMM sign language recognizer.

The advantages of the proposed system are: SVD still minimizes the feature vector for a particular gesture image and the features are not affected by scaling or rotation of gestures within an image which makes the system more flexible. Features generated using this technique makes the feature vector unique for a particular gesture. Moreover, posture recognition using IT2FHMM is also superior in several aspects. First, it is always statistically correct; unlike neural network. Second, it is easy to insert a new class of posture or to delete an existing class of posture and also new training for a specific posture is simply to accumulate new samples into its density function. This advantage is hard to achieve by other methods. Finally, because of the efficiency

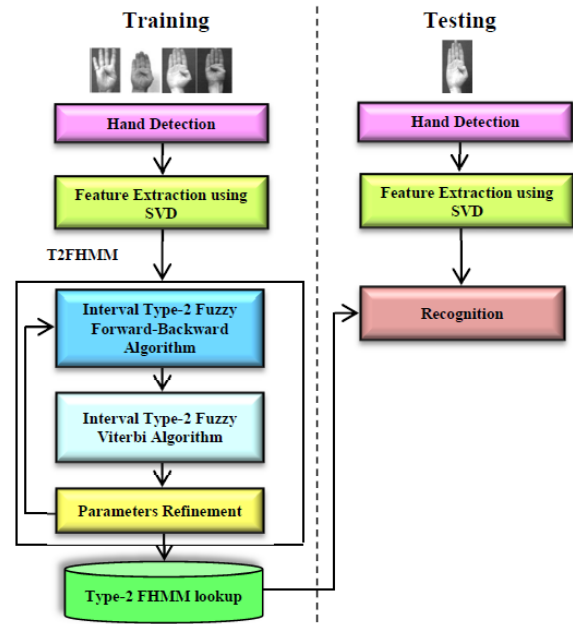in time and space, it is very easy to deal with large gesture "vocabularies".

Fig. 2. Proposed sign language recognizer.

A. Hand Detection

Hand detection is the key of success towards any gesture recognition due to challenges of vision based methods, such as varying lighting condition and complex background. Basically there are two methods for skin detection (i) Pixel-based segmentation and (ii) Region based segmentation. Pixel based segmentation uses color based methods [14]. The choice of the color model should consider separability between skin and non-skin color and should decrease separability among skin tones [14]. Color distribution has modeled into non parametric and parametric model. Non parametric model uses Histogram and Bays classifier for skin color distribution and parametric model derives Gaussian model from training set. Hand detection algorithm utilized here is simple, efficient and also easy to understand so that it can be applied to real time applications.

In general, Skin color based hand detection is unreliable for the difficulty to be distinguished from other skin-colored objects and sensitivity to lighting conditions. Approaches using shape models require sufficient contrast between object and background [6]. Considering tradeoff between computational cost and accuracy of detection, we use pixel based non parametric method under YCbCr color space [6]. The importance of YCbCr color model relies on two main aspects [15]; the Y component is separated from the Cb and Cr which are the chrominance components, and secondly these chrominance components depend on how human perceive this color spectrum. We ignore Y channel to reduce the effect of brightness variation and use only the chrominance channels that fully represent the color information. For more details about the conversion from RGB to YCbCr refer to [15].

This stage involves capturing the image through web camera having good quality. Here we are not keeping the background static and uniform. Light intensity remains uneven. So, segmentation requires lot of time to identify the area of interest.

B. Feature Extraction

Since hand gestures are very rich in shape variation, color, texture and motion, feature extraction is important for gesture recognition. Gestures are usually represented by various features, which are templates, transformations, geometric, and non-geometric features [1]. Templates are the easiest features to compute; they are simply the input gesture in its raw form without any extra calculations. Transformations, such as rotation, translation, or scaling can be applied to reduce the number of feature vectors in templates and then reduce the database size which may speed up the recognition time. Geometric features are the features that calculated somehow directly from the gesture skeleton or contour before or after applying some preprocessing operation; these features includes the width of the gesture, the height of the gesture; the number of fingers, the distance between the hand fingers, etc. ; however geometric features not always applicable due to self-occlusion and variations in lighting conditions. Non-geometric features include all the features that are extracted from applying some mathematical equations on a preprocessed hand gestures. The non-geometric features may be helpful in the identification of geometric features but are otherwise inadequate in recognition [16].

To increase the accuracy of the hand gesture recognition system and to overcome the problem of multi variations like rotation, scaling, translation, SVD feature selection approach is adopted to extract shape outline with minimum number of pixels for an image frame without losing shape information [4] in order to obtain more refined and distinguishable features. The common motivation for using SVD features extracted from raw data is dimensionality reduction, which would significantly reduce the size of the input vector.

The singular value decomposition, or SVD, is very powerful and useful matrix decomposition, particularly in the context of data analysis, dimension reducing transformations of data, and is the method of choice for solving most linear least–squares problems [8]. In formal, for any real $m \times n$ matrix A, there exist orthogonal matrices:

$$U = [u_1, u_2, ..., u_m] \in R^{m \times m},$$
$$V = [v_1, v_2, ..., v_n] \in R^{n \times n} \quad (1)$$

such that

$$A = U \sum V^T \quad (2)$$

where

$$\sum = diag\left(\sigma_1, \sigma_2, ..., \sigma_{\min(m, n)}\right) \in R^{m \times n},$$
$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min(m,n)} \geq 0 \quad (3)$$

σ_i is the ith singular value of A in non-increasing order, u_i and v_i are the ith left and right singular vectors of A for $i = \min(m, n)$, respectively. The singular values of a matrix A are unique, and the singular vectors corresponding to distinct singular values are uniquely determined up to the sign [8]. The ith largest singular value σ_i of A is actually the 2-norm or Euclidean length of the ith largest projected vector Ax onto x direction which is orthogonal to all the i -1 larger orthogonal vectors as shown by:

$$\sigma_i = \max_{U} \min_{x \in U, \|x\|_2 = 1} \|Ax\|_2 \quad (4)$$

where the maximum is taken over all i -dimensional subspaces $U \subseteq R^n$.

$$\sigma_1 = \max_{\|x\|_2 = 1} \|Ax\|_2 \quad (5)$$

C. Interval Type-2 Fuzzy Hidden Markov Model

A HMM represents a model of a stochastic process in which a set of observations is generated in discrete time by a sequence of states connected by transitions [11]. Depending on the type of observable sequences HMMs can be discrete or continuous. While the continuous speech recognition community adopted HMM's many years ago, these techniques are just now entering the vision community.

In this paper, we improve the HMMs expressive power for uncertainty by type-2 fuzzy set (T2FS). Uncertainties exist in both mode and data. A model may be interpreted as a set of elements and rules that map input variables onto output variables. Model uncertainty is uncertainty in the mapping induced by uncertain parameters in the model. Data uncertainty is uncertainty in the input variables [7]. The HMM characterizes these uncertainties by probability density function. Given sufficient training data, the HMM can accurately represent the training data according to the maximum likelihood (ML) criterion. In practice, however, the HMM generalize poorly to the test data because of noise, insufficient training data, and incomplete information. Therefore, modeling uncertainties is needed in both the HMM and hand gesture data.

The advantage of this extension is that it can handle both randomness and fuzziness within the framework of HMM by type-2 fuzzy sets (FSs). Membership functions (MFs) of type-2 fuzzy sets are three-dimensional that provides the additional degrees of freedom that make it possible to handle both uncertainties. Formally, T2FHMM is defined as follows [7]: $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_N\}$ is set of hidden fuzzy N states (total no. of classes in our case). $\tilde{q}_t$ encodes fuzzy state visited at time t, $2 \leq t \leq T$, where T is the total time, $\tilde{a}_{ij}$ denotes fuzzy transition probability from state $\tilde{S}_i$ to $\tilde{S}_j$, $\tilde{b}_j(o_t)$ is the observation o_t at time t's membership to the fuzzy state $\tilde{S}_j$, $h_{\tilde{\Omega}_t}(o_t)$ symbols T2MF of the non-singleton fuzzified observation vector o_t, c_{jm} is weight of the mth mixture component in fuzzy state $\tilde{S}_j$, μ_{jm} characterizes vector of means for the mth mixture component of fuzzy state $\tilde{S}_j$, $\sum_{jm}$ is covariance matrix for the mth mixture component of fuzzy state $\tilde{S}_j$, and $\tilde{\lambda}$ determines the set of all parameters defining a T2FHMM.

Inside T2FHMM, forward-backward or Baum–Welch algorithm and Viterbi search algorithm are efficient algorithms for HMM based training and recognition. Given

a set of training observation sequences, the Baum-Welch algorithm can iteratively and automatically adjust parameters μ and $\sum$ in the m th mixture component of the HMM. The Baum–Welch algorithm, an implementation of the expectation-maximization (EM) algorithm, guarantees that the model converges to a local maximum likelihood (ML) of observations of the training set. Using the Viterbi algorithm, we can decode the maximum likelihood state sequence $S^* = \arg\max_S P(O, S \mid \lambda)$ given an observation sequence O and an HMM λ, where the best state sequence S^* may represent signs in hand gesture recognition.

General type-2 FLS is computationally intensive because type-reduction is very intensive. Things simplify a lot when secondary membership functions (MFs) are interval sets (in this case, the secondary memberships are either zero or one and we call them interval type-2 sets).

1) Interval type-2 fuzzy baum–welch algorithm

Each IT2HMM can be trained using the Baum-Welch algorithm, applied to a set of images with hand gestures from the database, one set (20 images) for each gesture. For each training set we obtain a set of observable sequences, each set being subsequently used in training a HMM. The first step in T2FHMM training consists in initializing parameters as follows:

$$\tilde{a}_j(t) = [\underline{a}_j(t), \overline{a}_j(t)], \quad \underline{a}_1(1) = \overline{a}_1(1) = 1 \tag{6}$$

$$\underline{a}_j(1) = \underline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\underline{h}_{\tilde{\Omega}}(o_1) \times \underline{b}_j(o_1) \right) \right\rfloor$$

$$2 \leq j \leq N-1 \tag{7}$$

$$\overline{a}_j(1) = \overline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\overline{h}_{\tilde{\Omega}}(o_1) \times \overline{b}_j(o_1) \right) \right\rfloor,$$

$$2 \leq j \leq N-1. \tag{8}$$

Then the recursion process is started as follows

$$\underline{a}_j(t) = \left[\sum_{i=2}^{N-1} (\underline{a}_i(t-1) \times \underline{a}_{ij}) \right] \times \left\lfloor \sup_{o_t \in \tilde{\Omega}_t} \left(\underline{h}_{\tilde{\Omega}}(o_t) \times \underline{b}_j(o_t) \right) \right\rfloor \tag{9}$$

$$\overline{a}_j(t) = \left\lfloor \sum_{i=2}^{N-1} (\overline{a}_i(t-1) \times \overline{a}_{ij}) \right\rfloor \times \left\lfloor \sup_{o_t \in \tilde{\Omega}_t} \left(\overline{h}_{\tilde{\Omega}}(o_t) \times \overline{b}_j(o_t) \right) \right\rfloor \tag{10}$$

so, the final condition is computed by

$$\underline{a}_N(T) = \sum_{i=2}^{N-1} \left(\underline{a}_i(T) \times \underline{a}_{iN} \right), \quad \overline{a}_N(T) = \sum_{i=2}^{N-1} \left(\overline{a}_i(T) \times \overline{a}_{iN} \right) \tag{11}$$

The total membership grade $h_{\tilde{\lambda}}(O)$ is

$$h_{\tilde{\lambda}}(O) = [\underline{h}_{\tilde{\lambda}}(O), \overline{h}_{\tilde{\lambda}}(O)] \tag{12}$$

$$\underline{h}_{\tilde{\lambda}}(O) = \underline{a}_N(T), \quad \overline{h}_{\tilde{\lambda}}(O) = \overline{a}_N(T) \tag{13}$$

Regarding backward variable, in a similar way, we have

$$\beta_j(t) = [\underline{\beta}_j(t), \overline{\beta}_j(t)] \tag{14}$$

$$\underline{\beta}_i(T) = \underline{a}_{iN}, \overline{\beta}_i(T) = \overline{a}_{iN}, 2 \leq i \leq N-1 \tag{15}$$

$$\underline{\beta}_i(t) = \sum_{j=2}^{N-1} \left\{ \underline{a}_{ij} \times \left\lfloor \sup_{o_{t+1} \in \tilde{\Omega}_{t+1}} \left(\underline{h}_{\tilde{\Omega}_{t+1}}(o_{t+1}) \times \underline{b}_j(o_{t+1}) \right) \right\rfloor \right. \times \underline{\beta}_j(t+1) \left. \right\} \tag{16}$$

$$\overline{\beta}_i(t) = \sum_{j=2}^{N-1} \left\{ \overline{a}_{ij} \times \left\lfloor \sup_{o_{t+1} \in \tilde{\Omega}_{t+1}} \left(\overline{h}_{\tilde{\Omega}_{t+1}}(o_{t+1}) \times \overline{b}_j(o_{t+1}) \right) \right\rfloor \right. \times \overline{\beta}_j(t+1) \left. \right\} \tag{17}$$

$$\underline{\beta}_1(1) = \sum_{j=2}^{N-1} \left\{ \underline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\underline{h}_{\tilde{\Omega}_1}(o_1) \times \underline{b}_j(o_1) \right) \right\rfloor \times \underline{\beta}_j(1) \right\} \tag{18}$$

$$\overline{\beta}_1(1) = \sum_{j=2}^{N-1} \left\{ \overline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\overline{h}_{\tilde{\Omega}_1}(o_1) \times \overline{b}_j(o_1) \right) \right\rfloor \times \overline{\beta}_j(1) \right\} \tag{19}$$

Obviously, $\tilde{a}_j(t)$, $h_{\tilde{\lambda}}(O)$ and $\tilde{\beta}_j(t)$ are all IT1 sets.

$$\underline{a}_j(t) \times \underline{\beta}_j(t) = \underline{h}_{\tilde{\lambda}}(O, \tilde{q}_t = \tilde{S}_j) \tag{20}$$

$$\overline{a}_j(t) \times \overline{\beta}_j(t) = \overline{h}_{\tilde{\lambda}}(O, \tilde{q}_t = \tilde{S}_j) \tag{21}$$

2) Interval type-2 fuzzy viterbi algorithm

After the model parameters were initialized, the non-uniform segmentation of the images from the training set is replaced by Viterbi segmentation and the model parameters are recalculated. This step is an iterative one and it ends when the probabilities of the Viterbi segmentation for two successive iterations are lower than a given threshold. The final parameters of the T2FHMM model are obtained using the recursive Baum-Welch algorithm. The Viterbi algorithm chooses the best state sequence that maximizes the likelihood of the state sequence for the given observation sequence [17]. The suggested system adapts the product t-norm and maximum t-conorm in the meet and joins operations. The maximum membership grade of the first t observations at state $\tilde{S}_j, \tilde{\phi}_j(t) = [\underline{\phi}_j(t), \overline{\phi}_j(t)]$, can be computed by the following recursion:

$$\underline{\varphi}_j(t) = \max_{2 \leq i \leq N-1} \left\{ \underline{\varphi}_i(t-1) \times \underline{a}_{ij} \right\} \times \left\lfloor \sup_{o_t \in \tilde{\Omega}_t} \left(\underline{h}_{\tilde{\Omega}_t}(o_t) \times \underline{b}_j(o_t) \right) \right\rfloor \tag{22}$$

$$\overline{\varphi}_j(t) = \max_{2 \leq i \leq N-1} \left\{ \overline{\varphi}_i(t-1) \times \overline{a}_{ij} \right\} \times \left\lfloor \sup_{o_t \in \tilde{\Omega}_t} \left(\overline{h}_{\tilde{\Omega}_t}(o_t) \times \overline{b}_j(o_t) \right) \right\rfloor \tag{23}$$

where

$$\underline{\phi}_1(1) = \overline{\phi}_1(1) = 1 \tag{24}$$

$$\underline{\varphi}_j(1) = \underline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\underline{h}_{\tilde{\Omega}_1}(o_1) \times \underline{b}_j(o_1) \right) \right\rfloor \tag{25}$$

$$\overline{\varphi}_j(1) = \overline{a}_{1j} \times \left\lfloor \sup_{o_1 \in \tilde{\Omega}_1} \left(\overline{h}_{\tilde{\Omega}_1}(o_1) \times \overline{b}_j(o_1) \right) \right\rfloor \tag{26}$$

The maximum membership grade $h_{\tilde{\lambda}}(\wp^*, O)$ along the best state sequence $\wp^*$ is then given by

$$h_{\tilde{\lambda}}(\wp^*, O) = [\underline{h}_{\tilde{\lambda}}(\wp^*, O), \overline{h}_{\tilde{\lambda}}(\wp^*, O)] \tag{27}$$

$$\underline{h}_{\tilde{\lambda}}(\wp^*, O) = \underline{\varphi}_N(T) = \max_{2 \leq i \leq N-1} \left\{ \underline{\varphi}_i(T) \times \underline{a}_{iN} \right\} \tag{28}$$

$$\overline{h}_{\tilde{\lambda}}(\wp^*, O) = \overline{\varphi}_N(T) = \max_{2 \leq i \leq N-1} \left\{ \overline{\varphi}_i(T) \times \overline{a}_{iN} \right\} \tag{29}$$

If we use the center of the interval as the defuzzified values, then

$$\wp^* = \arg \max_{\wp} \left(\frac{\underline{h}_{\tilde{\lambda}}(\wp, O) + \overline{h}_{\tilde{\lambda}}(\wp, O)}{2} \right) \qquad (30)$$

Similarly, $\tilde{\phi}_j(t)$ and $h_{\tilde{\lambda}}(\wp^*, O)$ are IT1 sets too. $\tilde{\phi}_j(t)$ represents the maximum membership grade of the first t observations o_1 to o_t and ends in state $\tilde{S}_j$.

3) Parameters refinement

Finally, we use the IT2 fuzzy forward-backward algorithm to refine the parameters of the initialized IT2 FHMM. The steps to perform parameters re-estimation are summarized as follows:

$$\hat{\mu}_{jm} = c \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_{jm}^r(t) o_t^r}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_{jm}^r(t)} \qquad (31)$$

$$\hat{\Sigma}_{jm} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_{jm}^r(t)(o_t^r - \hat{\mu}_{jm})(o_t^r - \hat{\mu}_{jm})'}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_{jm}^r(t)} \qquad (32)$$

$$\hat{c}_{jm} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_{jm}^r(t)}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} L_j^r(t)} \qquad (33)$$

$$L_{jm}^r(t) = \frac{1}{\xi(h_{\tilde{\lambda}}(O^r))} \xi\left(\tilde{a}_{jm}(t) \Pi \tilde{\beta}_{jm}(t) \right) \qquad (34)$$

$L_{jm}^r(t)$ is the defuzzified membership grade of o_t to the m mixture of state $\tilde{S}_j$. $\xi(A)$ denotes the defuzzified value of T1FS A, which is a mapping from a T1FS to a crisp scalar. The defuzzified membership grade of observation o_t being in state $\tilde{S}_j$. after that, calculate the IT2 fuzzy forward and backward variables of the IT2 non-singleton fuzzified observation sequence O^r, $1 \le r \le R$, for all states $\tilde{s}_j$, mixtures m, and times t. then for each state $\tilde{s}_j$, mixture m and time t, use the weight $L_{jm}^r(t)$ and the current observation o_t to update the accumulators for that mixtures. Use the final accumulator values to calculate new parameter values to produce a new IT2 FHMM. Finally if the average membership grade of all training observations Λ for this iteration is not higher than the value at the previous iteration then stop, otherwise repeat steps using the new IT2FHMM. Where R represents the total number of observations.

$$\Lambda = \left(\sum_{r=1}^{R} (\underline{h}_{\tilde{\lambda}}(\wp^*, O^r) + \overline{h}_{\tilde{\lambda}}(\wp^*, O^r)) / 2 \right) / R \qquad (35)$$

D. Recognition

The main job of the classifier or recognizer is to identify which trained class the current presented testing gesture belongs to. The recognizer has to score those trained classes according to the presented testing gesture and the maximum matching score in the corresponding trained gesture class [9]. The existence of an efficient recognition algorithm plays a major role toward the success of the gesture recognition system, this leads toward hard algorithms which means time consuming and then will conflict with the real time applications which the speed is very important. In other hand, gesture recognition algorithm should be fast to compute, this abates the accurateness of the recognition, so, depending on the application of the system; one has to choose the classification algorithm that meets the demand.

In our case, the hand gesture is recognized by selecting the maximal observation probability of gestures T2FHMM model. The maximal gesture model is the gesture whose observation probability is the largest among all gestures. The type of observed gesture is decided by Viterbi algorithm frame by frame (i.e. accumulatively until it receives the gesture end signal). The Viterbi algorithm is used for evaluation at recognition time. Note that since Viterbi only guarantees the maximum of probability over all state sequences instead of the sum over all possible state sequences, the resultant scores are only an approximation. The image is recognized as the hand gesture, whose model has the highest production probability [12].

To conclude, we have built a system that allows online training and learning of new gestures. As seen in Fig. 2, the typical procedure for the interactive, online training and learning is as follows: (1) The user makes a certain gesture; (2) The data stream is segmented and is given as an input to the HMM process, then it is classified as either a predefined gesture or an unknown gesture; (3) If the gesture is a known gesture, the system recognizes it and updates the database simultaneously. If it is an unknown gesture, the system will add this record to the database and the user can define it as a new HMM state. We have presented the integration of type-2 fuzzy logic to the Hidden Markov Models. This integration consists in replacing the basic arithmetic operations by the adequate fuzzy operators. Using the fuzzy operators permits us in one hand to manage the imprecision concerning the data, and in the other hand it relaxes the additively constraint, necessary for the HMM, towards the monotonicity one, much less restrictive [7].

IV. EXPERIMENTAL RESULTS

The accuracy and performance of the proposed IT2FHMM sign language recognizer are further verified using an experimental dataset consisting of singlet universal sign language letters. All the experiments are carried out using Matlab R2011a version (7.13.0.564). The proposed algorithms have been implemented on i5-2430M CPU @ 2.40GHz with 4GB RAMs supplied with Microsoft Windows 7 Ultimate operating system.

Fig. 3. Examples of images used in IT2FHMM model training.

The training dataset for letter recognition contains more than 100 images per hand sign (see Fig. 3), so totally $6 \times 100 = 600$ images. Those images are subjected to illumination changes, scaling, mirroring, blurring, rotation,

view-point variant and translated. The lighting condition and background are also different in all data set. The subject's hand (mainly the color, texture and nails) has gone through some changes. Resolution is 56×46 for all the images that are normalized, smoothed before processing and converted into gray scale. The images used in testing were not used during the training, and their background is not similar to that of the training images. The utilized database was downloaded from the site http://www.idiap.ch/resource/gestures/. SVD feature descriptors are calculated for all the key points obtained for all the image samples in the database. The employed number of (feature vector) coefficients is 52.

The first set of experiments measures the efficiency of the proposed method by comparing it with other existing fuzzy HMM (FHMM) signer recognition method. (see Table I) reports the comparison between the two systems using the same features. The recognition accuracy was defined as the ratio of number of gestures correctly recognized to the total number of gestures tested. We can see that the IT2 FHMM has comparable classification ability to the FHMM in all data set. The overall success rate is 95.5% with the recognition time about 0.064 seconds. In other worlds, the proposed system offers promising potential to solve difficult hand gestures recognition. Experiment shows that the system can work correctly with sufficient training data, and recognition accuracy increases when the system is trained with more samples (see Table I- TableIV).

TABLE I: COMPARISON OF PROCESSING TIME AND RECOGNITION RATE

Class	FHMM		IT2FHMM	
	Time (Sec)	Average Recognition Rate (%)	Time (Sec)	Average Recognition Rate (%)
A	0.045	68.82	0.063	98.2
B	0.05	60	0.07	82.2
C	0.055	100	0.07	100
Five	0.048	66.25	0.064	97.5
Point	0.05	66.25	0.06	100
V	0.05	68.82	0.06	95.11
Overall average	**0.049**	**71.69**	**0.064**	**95.5**

TABLE II: CONFUSION MATRIX (10 SIGNER SAMPLES PER CLASS)

Input samples	A	B	C	Five	Point	V	Accuracy
A	**9**	0	1	0	0	0	90%
B	0	**7**	3	0	0	0	70%
C	0	0	**10**	0	0	0	100%
Five	0	0	1	**9**	0	0	90%
Point	0	0	0	0	**10**	0	100%
V	0	0	1	0	0	**9**	90%
Accuracy	90%	70%	100%	90%	100%	90%	90%

TABLE III: RECOGNITION RATE OF DIFFERENT SIZE OF SVD FEATURES

IT2FHMM	No. of SVD features (Coefficients)		
	20 features	30 features	50 features
Overall Recognition Rate	91.3	92.6	**95.5**

SVD for hand posture recognition alone takes little more processing time for an image because of the stack of computations formed by the algorithm. But it is robust against rotation, translation, scaling and illumination changes. For each key point 1x52 feature vector is obtained. SVD method with IT2FHMM recognizer shows 82.2% and

95.11% recognition rate with B and V set respectively with processing time of more than 0.06 sec. SVD features with FHMM recognizer has shown lower performance with recognition rate of 60% for B set. It is 1.5 times faster than IT2FHMM recognizer, because the computational complexity of the type-2 FHMM is higher than of FHMM. But it is not robust against rotation and illumination changes. Recognition rate and speed of processing of the proposed method are found to be promising. But shows poor performance with the B set where the illumination change occurs in dramatically.

In common, the high recognition rate on the training data indicates that the IT2FHMM topologies are sound and that the models are converging. Even so, the remaining 4.5% error rate on the case indicates that some fine tuning on the feature set and model is in order. With a larger training set and context modeling, lower error rates are expected and generalization to a freer, person-independent sign language recognition system should be attainable.

In experiment 2, the accuracy of obtained IT2FHMM recognizer can be assessed by a confusion matrix opposing assigned class (column) of the samples by the classifier with their true original class (row) (see Table II). Entries along the main diagonal are correct recognitions. Entries other than those on the main diagonal are recognition errors. The obtained test results exhibits accurate sign language recognition with low false alarms; thus, showing the robustness of the proposed system.

In experiment 3 the influence of SVD feature vector lengths 20, 30, and 50 in feature extraction phase for the recognition was verified (See Table III) that demonstrates the recognition rate as a function of the SVD vector length. As expected, the recognition rate increases slowly by increasing the number of features. Obviously, adaptive determination of a proper feature size is of great importance.

TABLE IV: PRECISION FOR TEST SET (THE VALUE IS PERCENTAGES)

Class		Rank3	Rank 4	Rank 5	Rank 6
A	FHMM	0.2	0.6	0.73	0.8
	IT2FHMM	1	1	1	1
B	FHMM	0.2	0.6	0.73	0.8
	IT2FHMM	1	1	1	1
C	FHMM	0.8	0.4	0.26	0.2
	IT2FHMM	1	1	1	1
Five	FHMM	0.2	0.6	0.73	0.8
	IT2FHMM	1	1	1	1
Point	FHMM	0.2	0.6	0.73	0.8
	IT2FHMM	1	1	1	1
V	FHMM	0.2	0.6	0.73	0.8
	IT2FHMM	1	1	1	1

In the last set of experiments, the precision for each test class are given in Table IV. The entries in column "Rank 3", for instance, is the percentage of cases that the correct result is ranked within the first three. As can be seen, the proposed type-2 fuzzy HMM system improves the precision performance with respect to fuzzy HMM for all chosen rank. This phenomenon is due to the fact that using HMM results in features spaces with type-2 membership function representation to handle high level of uncertainties associated with description of the hand and hence, most hand image lies in its relevant cluster in the space. Our system has an average precision of above 100 % vs. the top 3 images which means that our system has a good robustness to the sign categories.

Regarding system complexity, the computational complexity of the classical Forward- Backward procedure is $O(NS^2)$, where N is the number of states, and S is the total number of hand images for each gesture. If we choose bounded sum *t*-conorm in the *join* operation and the product *t*-norm in the *meet* operation then the type-2 fuzzy Forward- Backward procedure can be viewed as computing two boundary HMMs: the "lower" HMM and the "upper" HMM. Therefore, the computational complexity of the type-2 FHMM is twice that of the HMM.

V. CONCLUSIONS

Attempts at machine sign language recognition have begun to appear in the literature over the past ten years. We present a novel approach based on higher order fuzzy HMM to increase the recognition accuracy. Our method of gesture recognition is a vision based technique that does not use motion sensor gloves or colored gloves for the system to recognize hand shapes. T2FHMM-based framework is used for robust estimation the individual recognition units from the obtained SVD feature sequence. The set of images is trained by forward-backward algorithm then we create the database for matching the input image to the system by Viterbi algorithm to choose the maximum likelihood of the gestures. A sequence of recognition units is interpreted as a meaningful gesture.

The system is fully automatic and it works in real-time. It is fairly robust to background easy to use cluster, the system lies in the ease of its use. The proposed method's results show the potential of the system to correctly recognize a large variety of the signs no matter what is orientation of the hand in a particular sign. Type2 Fuzzy logic caters for the natural fuzziness involved in the problem of sign language and gives a very high accuracy rate.

Experiments on a single hand database have been carried out and recognition accuracy of up to 96% has been achieved with various hand gesture posture aspect ratios and under complicated backgrounds. Comparative study of proposed system with other popular techniques shows that the robustness is better. The number of samples/class represents a trade-off between the accuracy of recognizing the gesture and the amount of special/temporal data. Our future orientation concerns the use of other hand feature extraction techniques. We also think about identifying other large number of signs.

REFERENCES

[1] M. M. Hasan, and P. K. Mishra, "Robust gesture recognition using gaussian distribution for features fitting," *International Journal of Machine Learning and Computing*, vol. 2, no. 3, pp. 266-273, June 2012.

[2] S. Kausar, M. Y. Javed, and S. Sohail, "Recognition of Gestures in Pakistani Sign Language using Fuzzy Classifier," in *Proc. the 8th International Conference on Signal Processing, Computational Geometry and Artificial Vision*, Greece, 2008, pp. 101-105.

[3] A. B. Jmaa, W. Mahdi, Y. B. Jemaa, and A. B. Hmadou, "A new approach for digit recognition based on hand gesture analysis," *International Journal of Computer Science and Information Security*, vol. 2, no. 1, pp. 1-8, 2009.

[4] J. Rekha, J. Bhattacharya, and S. Majumder, "Hand gesture recognition for sign language: A new hybrid approach," in *Proc. the Conference on Image Processing Computer Vision and Pattern Recognition,* pp. 1-7, USA, January 2011.

[5] K. S. Patwardhan and S. D. Roy, "Dynamic hand gesture recognition using predictive eigen tracker," in *Proc. the Fourth Indian Conference on Computer Vision, Graphics & Image Processing*, 2004, pp. 1-6.

[6] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method," in *Proc. the IEEE International Conference on Multimedia and Expo*, 2007, pp. 995–998.

[7] J. Zeng and Z. Q. Liu, "Type-2 fuzzy hidden markov models and their application to speech recognition," *IEEE Transaction on Fuzzy Systems*, vol. 14, no. 3, pp. 454 – 467, June 2006.

[8] K. P. Kumar, P. Prasad, M. V. Ramakrishna, and B. Prasad, "Feature extraction using sparse SVD for biometric fusion in multimodal authentication," *International Journal of Network Security & Its Applications*, vol. 5, no. 4, pp. 83–94, July 2013.

[9] P. V. Kishore and P. R. Kumar, "A video based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic," *International Journal of Engineering and Technology*, vol. 4, no. 5, pp. 537–542, October 2012.

[10] M. Su, "A fuzzy rule-based approach to spatio-temporal hand gesture recognition," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 30, no. 2, pp. 276–281, May 2000.

[11] V. Tataru, R. L. Vieriu, and L. Goras, "On hand gestures recognition using hidden markov models," *Electronics and Telecommunications*, vol. 51, no. 3, pp. 29-32, 2010.

[12] N. D. Binh, E. Shuichi, and T. Ejima, "Real-time hand tracking and gesture recognition system," in *Proc. the Conference on Graphics, Vision and Image Processing*, Cairo, Egypt, 2005, pp. 362–368.

[13] H. Wang, M. C. Leu, and C. Oz, "American sign language recognition using multi-dimensional hidden markov models," *Journal of Information Science and Engineering*, vol. 22, no. 5, pp. 1109–1123, 2006.

[14] A. S. Ghotkar and G. K. Kharate, "Hand segmentation techniques to hand gesture recognition for natural human computer interaction," *International Journal of Human Computer Interaction*, vol. 3, no. 1, pp. 15–25, 2012.

[15] N. A. Ibraheem, M. M. Hasan, R. Z. Khan, and P. K. Mishra, "Understanding color models: a review," *Journal of Science and Technology*, vol. 2, no. 3, pp. 265–275, April 2012.

[16] R. Verma and A. Dev, "Vision based hand gesture recognition using finite state machines and fuzzy logic," *International Conference on Ultra-Modern Telecommunications and Workshops*, pp. 1–6, Russia, 12-14 October 2009.

[17] M. Bansal, S. Saxena, D. Desale, and D. Jadhav, "Dynamic gesture recognition using hidden markov model in static background," *International Journal of Computer Science Issues*, vol. 8, no. 6, pp. 391-398, November 2011.

Saad M. Darwish received his Ph.D. degree from the Alexandria University, Egypt. His research and professional interests include image processing, optimization techniques, security technologies, and machine learning. He has publications in journals and conferences and severed as a TPC of many international conferences. Since Feb. 2012, he has been an associate professor in the Department of Information Technology, Institute of Graduate Studies and Research, Egypt.

Magda M. Madbouly received her Ph.D. degree from the Alexandria University, Egypt. Her research and professional interests include artificial intelligence, cloud computing, neural networks and machine learning. She is an assistant professor in the Department of Information Technology, Institute of Graduate Studies and Research, Egypt.

Murad B. Khorsheed received the B.Sc. degree in computer sciences from the Faculty of Sciences, University of Kirkuk, Iraq in 2007. Currently he is a M.Sc. student in the Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, Egypt. His research and professional interests include image processing and machine learning.