

MAN - MACHINE INTERACTION SYSTEM FOR SUBJECT INDEPENDENT SIGN LANGUAGE RECOGNITION USING FUZZY HIDDEN MARKOV MODEL

SAAD M. DARWISH

ABSTRACT. Sign language recognition has spawned more and more interest in humancomputer interaction society. The major challenge that SLR recognition faces now is developing methods that will scale well with increasing vocabulary size with a limited set of training data for the signer independent application. The automatic SLR based on hidden Markov models (HMMs) is very sensitive to gesture's shape information that makes the accurate parameters of the HMM not capable of characterizing the ambiguous distributions of the observations in gesture's features. This paper presents an extension of the HMMs using interval type-2 fuzzy sets (IT2FSs) to produce interval type-2 fuzzy HMMs to model uncertainties of hypothesis spaces (unknown varieties of parameters of the decision function). The benefit of this enlargement is that it can control both the randomness and fuzziness of traditional HMM mapping. Membership function (MF) of type-2 FS is three-dimensional that provides additional degrees of freedom to evaluate HMM's uncertainties. This system aspires to be a solution to the scalability problem, i.e. has real potential for application on a large vocabulary. Furthermore, it does not rely on the use of data gloves or other means as input devices, and operates in isolated signer-independent modes. Experimental results show that the interval type-2 fuzzy HMM has a comparable performance as that of the fuzzy HMM but is more robust to the gesture variation, while it retains almost the same computational complexity as that of the FHMM.

1. Introduction

The purpose of the Sign language recognition (SLR) is to offer a precise and suitable tool to set down sign gestures into expressive text or speech so that interaction between deaf and hearing people can effortlessly be made [22]. Hand gestures are spatio-temporally altering and therefore the automatic gesture recognition arises to be exceedingly stimulating. The key challenge in sign language recognition is to discover an adequate model to grab hold of the language, yet scales to outsized words [27, 24]. Gesture Recognition can be realized by one of the two procedures, one, by tiring sensual gloves on ones hand or second, by with the assistance of computer vision [8]. For glove based methods, it essentially exploits sensory gloves to quantify the angles and spatial locations of a hand and fingers. For computer

Received: ****; Revised: ****; Accepted: ***

Key words and phrases: Hand Gesture Recognition, Sign Language, Type-2 Fuzzy Logic, Hidden Markov Model.

vision- based methods, one or a set of cameras are involved in acquiring images for hand gesture recognition. The employment of computer-generated gloves or other resources of input devices disagreements with identifying gestures in a usual perspective and is very hard to implement in real time. So, recently, researchers offered numerous SLR systems based on computer vision techniques [2, 4]. Normal humancomputer interaction (HCI) should be glove free, fast, dependable and suitable.

Most of the widespread hand communicating systems can be investigated to be encompassed of three layers [15, 23]: detection, tracking, and recognition. The detection layer is accountable for determining and eliciting visual features that can be accredited to the occurrence of hands in the field of view of the camera(s). The tracking layer is responsible for carrying out temporal data relationship among sequential image frames; so that, at each instant in time, the system may be informed of "what is where". While the recognition layer is in charge for merging the spatiotemporal data taken away in the preceding layers and conveying the resultant clusters with labels related to specific classes of gestures. Sign language recognition can be characterized into isolated and continuous SLR and each can be auxiliary categorized into signer-dependent and signer-independent based on the reactivity to the signer. Signer independence is extremely required since it permits a system to be applied directly to the box and it lets the system be constructed for the signer who is not recognized earlier [17].

For signer-independent SLR, there are two problems [17, 1]: (1) the model convergence complexity triggered by perceptible dissimilarities between diverse person signs. For a powerful recognition model, the training data must be composed of numerous signers. This causes the training data very immense. (2) The lack of useful features extorted from several signers data. Unlike speech recognition in which every speech signature has been intensely discovered, the research on the feature extraction of SLR is still in its beginning. How to successfully extract public features from different signers is a more stimulating problem that requires being solved.

Different approaches have been exploited for signer-independent SLR, among them HMM has been the commonly used techniques due to its advantages [1, 10, 29, 36]. HMM has the facility to model a time domain process when taking into account the positions and orientations of gestures across time. But, the traditional HMM has some restrictions. One of them is the assumption that the allocations of distinct observation parameters can be well characterized as a mixture of Gaussian. Another significant matter of HMMs for gesture recognition is the scalability. The SLR vocabulary covers thousands of signs. Training and testing thousands of HMMs are very hard and not realistic. In general, the above two problems of signer-independent SLR lead to the fact that sign representations cannot be well modeled by conventional HMM.

In the literature, the kinds of vagueness in gesture recognition may be certain fuzziness and non-specificity causing from incomplete information, which contains fuzzy decision functions (uncertain mapping or uncertain hypothesis space), fuzzy observations (noise or non-stationary data), and fuzzy similarity match (uncertain

matching degree) [16, 6, 5, 14]. The HMM is completely certain once its parameters are stated. Still, those parameters may not precisely reveal the fundamental distribution of the observations because of inadequate or noisy data in sign recognition problems. In addition, it may appear difficult to hire likelihoods that are exact real numbers to assess uncertain HMMs with respect to the observation. Although this does not establish a critical problem for many applications, it is yet feasible to express the uncertain parameters of the HMM to permit for the uncertain likelihoods. A practical approach is to use the generalized HMM with fuzzy measure and fuzzy integral (Fuzzy hidden Markov model) [5]. In recent years, many academics have engaged type 2 Fuzzy HMM for sign language recognition and achieved inspirational results [5] [34, 3, 25]. The advantage of this extension is that it can handle both randomness and fuzziness within the framework of type-2 fuzzy sets. Membership functions of type-2 fuzzy sets are three-dimensional that provides the additional degrees of freedom that make it possible to handle both uncertainties.

Yet, type-2 fuzzy sets for sign language recognition are complicated to recognize and use because: (1) the three-dimensional nature of type-2 fuzzy sets makes them very hard to draw; (2) derivations of the formulas for the union, intersection, and complement of type-2 fuzzy sets all depend on Zadehs extension principle, which in itself is a difficult concept, (3) using type-2 fuzzy sets is computationally more complicated than using type-1 fuzzy sets, and (4) Performance of a sign recognition system significantly depends on feature selection and classifier design. A good classification algorithm occasionally cannot produce high classification accuracy for out of sorts selected features. On the other hand, even using a large set of features, describing a gesture, we irregularly fail to distinguish the gesture properly because of a poor classifier [22, 5, 14].

1.1. Motivation. The realization of interval type-2 fuzzy sets has been essentially qualified to their three-dimensional membership functions to manage more uncertainties in real-world problems. In gesture recognition, both feature and hypothesis spaces have uncertainties. In the feature space, arbitrary observations are usually extracted by the class-conditional probability density functions. In the hypothesis space, the factors of the decision function are random variables with some recognized prior distributions, and training data change this distribution of the variables into posterior probability density. Given sufficient training data, the HMM can precisely characterize the training data. In practice, however, the HMM generalizes poorly to the test data because of noise, insufficient training data, and incomplete information. Therefore, modeling uncertainties are needed in HMM. This motivates us for integrating interval type-2 fuzzy sets with HMM classifiers to achieve a better performance in terms of the robustness, generalization ability, or recognition accuracy.

1.2. Contribution and Novelty. Motivated by the desire to deliver users with an instinctive gesture input system that has good performance on a large vocabulary size with a limited set of training data; the work presented in this paper designates a novel type-2 fuzzy Hidden Markov Model (T2FHMM) based framework to tackle the problems of signer-independent SLR. The recommended visionbased system will

be intended to maximize the recognition ratio for gesture database under unconstrained environments in case of illumination changes, scaling, and rotation. This system is more appropriate for larger vocabulary as computation time and training data size; do not grow extremely with the vocabulary size as in most previous methods that deal with gloves-based signer dependent applications.

Despite its applications to a mass of decision functions' tasks, according to recent reviews [6, 14], interval type-2 fuzzy logic in conjunction with HMM classifier has not been adapted to SLR tasks, and in this respect, the present article represents a novel methodology. The system recognizes a gesture based upon the sequence of hand features to model the various changes of the hand postures. The observation sequence used to describe the states of the IT2FHMM are acquired from the features extracted from the segmented hand image through Singular Value Decomposition (SVD), which is very powerful and useful matrix decomposition, especially in the context of data analysis and dimension reducing transformations [18].

The rest of the paper is organized as follows. Section 2 describes some of the recent related HMM-based works. Section 3 describes the proposed system. The test results and discussion of the meaning are shown in section 4. A short summary of this paper and outlook of future work are given in section 5.

2. Related Work

In general, hand gesture recognition can be categorized into two methods [22] [20, 21, 11]. (a) Rule-based approaches that consist of a set of manually coded rules among feature inputs. Given an input gesture, a set of features are extracted and matched to the encoded rules; the rule that balances the input is outputted as the gesture. The main difficulty with rule-based approaches is that they depend on the skill of a human to encode rules. (b) Machine learning based approaches that handle a gesture as the output of a stochastic process. Of this category of approach, HMMs [1, 10, 29] by far have obtained the most consideration in the literature for categorizing gestures. For details on HMM-based gesture recognition systems, the survey in Reference [21] contributed a full review. However, there are still several challenges for correct and real-time SLR with large vocabulary signs.

In fact, the method of combining artificial neural network (ANN) with HMM [11] is an ideal alternative to overcome HMMs limitations such as inability to discriminate training data in general. The hybrid paradigm maintains an underlying HMM structure, capable of modeling long term dependencies, with the integration of ANN, which provides a probability estimation, discriminative training algorithms, and fewer parameters to estimate than those usually required in conventional HMM. However, there are still many challenges for accurate and real-time SLR with large vocabulary signs. The hybrid paradigm cannot adapt well itself to scale with the increasing of vocabulary size.

In previous work [7], researchers have proposed other extensions to HMMs to model the interaction of several interacting processes in parallel, such as factorial hidden Markov models or coupled hidden Markov models for recognizing complex gestures. These extensions require modeling the interactions of the processes during

the training phase, and thus require training examples of every conceivable combination of actions that can occur in parallel. Recently, the authors in [16] presented a fuzzy logic combination to the HMM. They have substituted the elementary arithmetic operators by some tolerable fuzzy operators. Using fuzzy operators allows to decrease the additivity restriction of probability measures. In general, the use of fuzzy operators makes it possible to process fuzzy data and to handle uncertainty, which is inherent in the HMM usage.

In light of the T2FS framework, many authors reconsider the HMMs vagueness and suggested a novel extension of the HMM. T2FS is likely to handle two kinds of uncertainties, namely, randomness and fuzziness; probability theory is related to the former and FS theory is accompanying with the latter [5]. Compared with the HMM, the T2FHMM manages a chain of T2 fuzzy vectors rather than a crisp numeric chain, and the output of the T2FHMM is an uncertain T1FS rather than a crisp scalar. In addition, type-2 FHMMs do not need additional training data and the computational complexity is almost the same as that of the HMM. Still, some of the main difficulties with type-2 fuzzy sets are the obstacle of understanding, imagining how they inspect due to their 3-D nature, and the computational complexity required to create solutions. By utilizing interval type-2 fuzzy sets, described by secondary membership functions capturing values of either 0 or 1, the type reduction required for defuzzification of type-2 fuzzy sets is simplified. Readers looking for more information regarding the comparison between HMM, FHMM, and T2FHMM can refer to [34]. Thus, it is doubtful that these extensions will scale well in SLR; but this is at the expense of complexity and processing time. However, most of the above systems are signer dependent SLR.

To the best of my knowledge, based on Google Scholar there has been very limited research work related to independent isolated SLR to recognize small vocabulary signs. For instance, the system in [10] used self-organizing feature map (SOFM) as a feature extractor for discrete HMM to recognize signer-independent Chinese sign language over the 4368 samples from 7 signers with 208 isolated signs. The other research in [11] reported a signer-independent system to distinguish a set of 52 signs. Their system engaged a modular architecture consisting of multiple feature-recognition neural networks and the nearest neighbor classifier to identify isolated signs. Still, these systems used "data gloves" as input devices for the recognition of SLR and suffer from limiting the user's freedom of movement. Video-based techniques are less intrusive and therefore more comfortable to utilize.

For video-based signer independent SLR techniques with a large vocabulary, even no research report was found in the literature. The work in [10] introduced an automatic Arabic SLR system based on HMM where a large set of samples has been used to recognize a few set of isolated words in a controlled lighting condition. The data collection phase imposed many restrictions on the background and illumination. Fuzzy logic is one of the artificial intelligence methods helps to solve the problem when there is gesture recognition with uncertainty and vagueness in it [25]. The main intention of this paper is to study about how effectively higher order fuzzy HMM can be used to improve the scalability of the SLR systems in

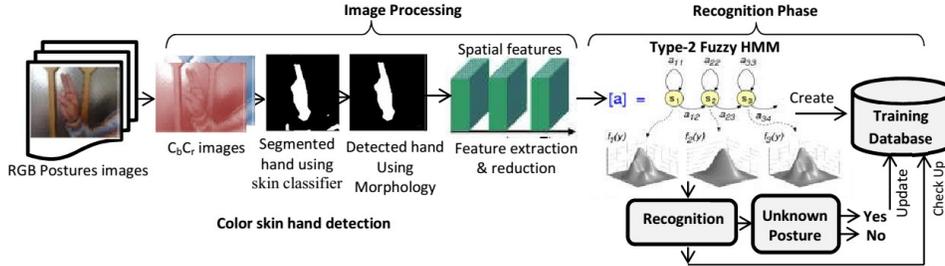


FIGURE 1. Block diagram of the proposed IT2FMMM-signer independent SLR

terms of signer independent application so that more hand signs can be processed and accurate model can be predicated.

3. Proposed IT2FHMM Sign Language Recognition System

The proposed postures recognition system is a vision based technique to identify isolated signs from images where no restrictions are imposed on the signer and the background with a sufficiently large vocabulary size to make them suitable for practical deployment. This system improves the HMMs classifier power for uncertainty by type-2 fuzzy set (T2FS). Membership functions (MFs) of type-2 fuzzy sets are three-dimensional in which the primary membership is used to describe the randomness in the parameters of the HMM while the secondary MF is employed to describe the fuzziness of the primary MF. Figure 1 shows the block diagram of the proposed HCI recognizer that is described below. To model postures signals, the left-right non-skip HMM topology is often used where no transitions are allowed to states whose indices are lower than the current state. The parameters of the T2FHMMs are defined with Fuzzy Gaussian distributions (uncertain mean) based on the training samples. For training, the type-2 fuzzy Baum-Welch algorithm is applied [5]. In the recognition phase, the Type-2 fuzzy Viterbi algorithm is used to find the most probable word sequence [34].

The advantages of the proposed system are: SVD still minimizes the feature vector for a particular posture image and the features are not affected by scaling or rotation of gestures within an image, which makes the system more flexible. Furthermore, features generated using this technique makes the feature vector unique for a particular gesture. The system can perform interactive training and online recognition of SLR postures in real time. It can also intelligently learn new postures, which can be used in recognition afterward.

Step 1: Hand Detection

Hand detection is the fundamental to success towards any gesture recognition due to challenges of vision-based methods, such as changing lighting circumstance and compound background. Basically, there are two methods for skin detection:

pixel (color) and region based segmentation [12, 30]. Skin color detection is undependable for the trouble to be famed from other skin-colored objects and sensitivity to lighting environments. Approaches using shape models (region based) need enough contrast between object and background. Bearing in mind the tradeoff between computational cost and accuracy of detection, the suggested system uses pixel based non-parametric (histogram) segmentation method under YCbCr color space. The color space transformation is supposed to decrease the overlap between the skin and non-skin pixels to categorize skin pixel and offer robust parameter against varying illumination conditions [13].

In this method, the input which is an image or a frame from a video can be acquired from web camera or any other camera. After rescaling the input image to diminish processing time; this RGB color image is transformed into CbCr chrominance image (chrominance vector). The system disregards the Y channel to decrease the effect of brightness variation and employs only the chrominance channels that completely characterize the color information. The final goal of skin color detection is to construct a decision rule that will distinguish between the skin and non-skin pixels. One method to produce a skin classifier is to outline obviously (through a number of rules) the boundaries skin cluster in CbCr color space. Refer to [30] for a comprehensive overview of skin cluster.

Step 2: Feature Extraction

Selecting features are vital to gesture recognition since hand gestures are rich in shape variation, motion, and textures. In general, features can be originated with the following three approaches: model (Kinematic Model), view, and low-level features (appearance) based methodologies [21, 28]. Provided the binary segmented hand image from the earlier step, to rise the correctness of the hand gesture recognition system and to overwhelmed the problem of multi-variations like rotation, scaling, and translation SVD feature selection approach is approved to extort shape skeleton with lowest number of pixels for an image frame without dropping shape information in order to find more distinguished and distinct features[21].The common stimulus for utilizing SVD features obtained from raw data is dimensionality reduction, which would meaningfully decrease the size of the input vector. Informal, for any real $m \times n$ matrix A , there exist orthogonal matrices [19]:

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

such that

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

where

$$\begin{aligned} \sum &= \text{diag} (\sigma_1, \sigma_2, \dots, \sigma_{\min(m, n)}) \in R^{m \times n}, \\ \sigma_1 &\geq \sigma_2 \geq \dots \geq \sigma_{\min(m, n)} \geq 0 \end{aligned} \quad (3)$$

σ_i is the i^{th} singular value of A in non-increasing order, u_i and v_i are the i^{th} left and right singular vectors of A for $i = \min(m, n)$, in that order. The i^{th} largest singular value σ_i of A is really the Euclidean length of the i^{th} largest projected

vector Ax onto x direction which is orthogonal to all the $i - 1$ larger orthogonal vectors as shown by:

$$\begin{aligned}\sigma_i &= \max_U \min_{x \in U, \|x\|_2=1} \|Ax\|_2, \\ \sigma_1 &= \max_{\|x\|_2=1} \|Ax\|_2\end{aligned}\quad (4)$$

where the maximum is occupied over all i -dimensional subspaces $U \subseteq R^n$. In this case, the HMM observable vectors consist of $\sigma_1, \sigma_2, \dots, \sigma_{\min(m, n)}$ coefficients stemmed by operating SVD transform for each training sign. The number and particularly the order of the coefficients play a significant role in producing acceptable recognition model.

Step 3: Type-2 Fuzzy Hidden Markov Model

There are three essential jobs in the HMM design in general: (1) Given an observation sequence, calculate the likelihood with which those observations can be created by a certain HMM model; (2) Regulate a most possible sequence of inner states in a particular model which will provide enlargement to a provided observation sequence; (3) Modify the model parameters of an HMM to enhance the possibility distribution matrices for a prearranged set of observations. The particulars of these tasks are explained in reference [31]. There are three key difficulties for HMM: evaluation, decoding, and training, which are resolved by exploiting Forward algorithm, Viterbi algorithm, and Baum-Welch algorithm, correspondingly [1, 10, 29].

To check the HMM uncertainty (The parameters of the HMM are uncertain because of inadequate and despoiled training data that destroy the HMMs expressive control to assess the unseen test data), type2 fuzzy sets is emphasized. Formally, T2FHMM is stated as follows [5, 32] $\tilde{S} = \{ \tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_N \}$ is set of unknown fuzzy N states (total number of classes). \tilde{q}_t codes fuzzy state stayed at time t $2 \leq t \leq T$, where T is the total time, \tilde{a}_{ij} signifies fuzzy transition chance from state \tilde{S}_i to \tilde{S}_j , $\tilde{b}_j(o_t)$ is the observation o_t at time t membership to the fuzzy state \tilde{S}_j , $h_{\tilde{Q}_t}(o_t)$ symbols T2MF of the non-singleton fuzzified observation vector o_t , c_{jm} is the weight of the m^{th} mixture component in fuzzy state \tilde{S}_j , μ_{jm} characterizes vector of means for the m^{th} mixture component of fuzzy state \tilde{S}_j , Σ_{jm} is the covariance matrix for the m^{th} mixture component of fuzzy state \tilde{S}_j , and $\tilde{\lambda}$ determines the set of all parameters defining a T2FHMM. $\tilde{\lambda}$ contains uncertain information carried by $\tilde{b}_j(o_t)$ and \tilde{a}_{ij} (in order to relax one of the HMMs intrinsic restrictions that is the observation distribution can be well characterized). Once more, the primary MF of $\tilde{b}_j(o_t)$ is exploited to form randomness and the secondary MF is employed to assess the fuzziness of the model. Here, Gaussian MF with uncertain mean is handled for operation of $\tilde{b}_j(o_t)$, see [35, 26] for more details.

The membership grade of \tilde{a}_{ij} exposes the uncertainty of the transition probability from \tilde{S}_i to \tilde{S}_j . The primary membership of $\tilde{b}_j(o_t)$ is the membership that o_t fits the fuzzy hidden state \tilde{S}_j , and the secondary grade reveals the trust to this membership. The T2FS signifies the uncertainties by two essential ideas: secondary membership function and footprint of uncertainty (FOU). Holding the domain in the interval $[0,1]$, the secondary MF is a map of the membership (not just a point of value as

type-1 (T1) MF) at each value of the input primary variable. The combination of all secondary MF domains constitutes a constrained region FOU imitating the degree of uncertainty of the model. Yet, general type-2 FLSs are computationally exhaustive because type-reduction is very intensive. Things abridge 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). Therefore interval type-2 fuzzy HMM (IT2FHMM) is more suitable for practical use.

(1) **Training Phase**

Given observable sequences for each training set, IT2HMM can be prepared with IT2 fuzzy form of Baum-Welch algorithm (IT2 fuzzy forward and backward). The first step in IT2FHMM training consists in resetting parameters where all variables can symbolize by the intervals of their lower and upper MFs $[\underline{\mu} \ \bar{\mu}]$ that is governed by previous knowledge where $\underline{\mu} = \mu - k \times \sigma$ and $\bar{\mu} = \mu + k \times \sigma$, $k \in [0, 3]$ and k is called uncertainty factor the regulates the uncertainty of the model and adjusts FOU. Here $*$ is the product t -norm and \sum is the confined sum t -conorm in the meet and join operations to implement the IT2FMM. IT2 fuzzy forward variables can be defined as follows [5, 35, 26]:

$$\begin{aligned} \bar{\alpha}_j(t) &= [\underline{a}_j(t), \bar{a}_j(t)], \\ \underline{\alpha}_1(1) &= \bar{\alpha}_1(1) = 1 \end{aligned} \quad (5)$$

$$\underline{\alpha}_j(1) = \underline{a}_{1j} * \left[\sup_{o_1 \in \bar{\Omega}_1} (\underline{h}_{\bar{\Omega}}(o_1) * \underline{b}_j(o_1)) \right] \quad 2 \leq j \leq N - 1 \quad (6)$$

$$\bar{\alpha}_j(1) = \bar{a}_{1j} * \left[\sup_{o_1 \in \bar{\Omega}_1} (\bar{h}_{\bar{\Omega}}(o_1) * \bar{b}_j(o_1)) \right] \quad (7)$$

then the recursion process is started:

$$\underline{\alpha}_j(t) = \left[\sum_{i=2}^{N-1} (\underline{\alpha}_i(t-1) * \underline{a}_{ij}) \right] * \left[\sup_{o_t \in \bar{\Omega}_t} (\underline{h}_{\bar{\Omega}}(o_t) * \underline{b}_j(o_t)) \right] \quad (8)$$

$$\bar{\alpha}_j(t) = \left[\sum_{i=2}^{N-1} (\bar{\alpha}_i(t-1) * \bar{a}_{ij}) \right] * \left[\sup_{o_t \in \bar{\Omega}_t} (\bar{h}_{\bar{\Omega}}(o_t) * \bar{b}_j(o_t)) \right] \quad (9)$$

so, the final condition is computed by:

$$\begin{aligned} \underline{\alpha}_N(T) &= \sum_{i=2}^{N-1} (\underline{\alpha}_i(T) * \underline{a}_{iN}), \\ \bar{\alpha}_N(T) &= \sum_{i=2}^{N-1} (\bar{\alpha}_i(T) * \bar{a}_{iN}) \end{aligned} \quad (10)$$

the total membership grade $h_{\bar{\lambda}}(O)$ is

$$\begin{aligned} h_{\bar{\lambda}}(O) &= [h_{\bar{\lambda}}(O), \bar{h}_{\bar{\lambda}}(O)], \\ \underline{h}_{\bar{\lambda}}(O) &= \underline{a}_N(T), \quad \bar{h}_{\bar{\lambda}}(O) = \bar{a}_N(T) \end{aligned} \quad (11)$$

regarding IT2 fuzzy backward variable, in a similar way, we have

$$\beta_j(t) = [\underline{\beta}_j(t), \bar{\beta}_j(t)] \quad (12)$$

$$\underline{\beta}_i(T) = \underline{a}_{iN}, \quad \bar{\beta}_i(T) = \bar{a}_{iN}, \quad 2 \leq i \leq N-1 \quad (13)$$

$$\begin{aligned} \underline{\psi}_1 &= \left[\sup_{o_{t+1} \in \bar{\Omega}_{t+1}} \left(\underline{h}_{\bar{\Omega}_{t+1}}(o_{t+1}) * \underline{b}_j(o_{t+1}) \right) \right], \\ \underline{\beta}_i(t) &= \sum_{j=2}^{N-1} \{ \underline{a}_{ij} * \underline{\psi}_1 * \underline{\beta}_j(t+1) \} \end{aligned} \quad (14)$$

$$\begin{aligned} \bar{\psi}_1 &= \left[\sup_{o_{t+1} \in \bar{\Omega}_{t+1}} \left(\bar{h}_{\bar{\Omega}_{t+1}}(o_{t+1}) * \bar{b}_j(o_{t+1}) \right) \right], \\ \bar{\beta}_i(t) &= \sum_{j=2}^{N-1} \{ \bar{a}_{ij} * \bar{\psi}_1 * \bar{\beta}_j(t+1) \} \end{aligned} \quad (15)$$

$$\underline{\beta}_1(1) = \sum_{j=2}^{N-1} \{ \underline{a}_{1j} * \left[\sup_{o_1 \in \bar{\Omega}_1} \left(\underline{h}_{\bar{\Omega}_1}(o_1) * \underline{b}_j(o_1) \right) \right] * \underline{\beta}_j(1) \} \quad (16)$$

$$\bar{\beta}_1(1) = \sum_{j=2}^{N-1} \{ \bar{a}_{1j} * \left[\sup_{o_1 \in \bar{\Omega}_1} \left(\bar{h}_{\bar{\Omega}_1}(o_1) * \bar{b}_j(o_1) \right) \right] * \bar{\beta}_j(1) \} \quad (17)$$

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

$$\underline{\alpha}_j(t) * \underline{\beta}_j(t) = \underline{h}_{\bar{\lambda}}(O, \tilde{q}_t = \tilde{S}_j) \quad (18)$$

$$\bar{\alpha}_j(t) * \bar{\beta}_j(t) = \bar{h}_{\bar{\lambda}}(O, \tilde{q}_t = \tilde{S}_j) \quad (19)$$

After the model parameters were reset, the non-uniform segmentation of the images from the training set is substituted by Viterbi segmentation and the model parameters are re-estimated. This step is an iterative procedure and it finishes when the likelihoods of the Viterbi segmentation for two consecutive iterations are lower than a prearranged threshold. The ending parameters of the IT2FHMM model are gained by means of the recursive IT2 fuzzy Baum-Welch algorithm. The IT2 fuzzy Viterbi algorithm selects the greatest state sequence that make the most of the likelihood of the state sequence for the specified observation sequence [5]. The maximum membership grade of the first t observations at state \tilde{S}_j , $\tilde{\phi}_j(t) = [\underline{\phi}_j(t), \bar{\phi}_j(t)]$, can be computed by the following recursion:

$$\begin{aligned} \underline{\psi}_2 &= \left[\sup_{o_t \in \tilde{\Omega}_t} (\underline{h}_{\tilde{\Omega}_t}(o_t) * \underline{b}_j(o_t)) \right], \\ \underline{\phi}_j(t) &= \max_{2 \leq i \leq N-1} \left\{ \underline{\phi}_i(t-1) * \underline{a}_{ij} \right\} * \underline{\psi}_2 \end{aligned} \quad (20)$$

$$\begin{aligned} \bar{\psi}_2 &= \left[\sup_{o_t \in \tilde{\Omega}_t} (\bar{h}_{\tilde{\Omega}_t}(o_t) * \bar{b}_j(o_t)) \right], \\ \bar{\phi}_j(t) &= \max_{2 \leq i \leq N-1} \left\{ \bar{\phi}_i(t-1) * \bar{a}_{ij} \right\} * \bar{\psi}_2, \\ \underline{\phi}_1(1) &= \bar{\phi}_1(1) = 1 \end{aligned} \quad (21)$$

where

$$\begin{aligned} \underline{\phi}_j(1) &= \underline{a}_{1j} * \left[\sup_{o_1 \in \tilde{\Omega}_1} (\underline{h}_{\tilde{\Omega}_1}(o_1) * \underline{b}_j(o_1)) \right], \\ \bar{\phi}_j(1) &= \bar{a}_{1j} * \left[\sup_{o_1 \in \tilde{\Omega}_1} (\bar{h}_{\tilde{\Omega}_1}(o_1) * \bar{b}_j(o_1)) \right] \end{aligned} \quad (22)$$

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

$$h_{\tilde{\lambda}}(\wp^*, O) = [\underline{h}_{\tilde{\lambda}}(\wp^*, O), \bar{h}_{\tilde{\lambda}}(\wp^*, O)] \quad (23)$$

$$\underline{h}_{\tilde{\lambda}}(\wp^*, O) = \underline{\phi}_N(T) = \max_{2 \leq i \leq N-1} \left\{ \underline{\phi}_i(T) * \underline{a}_{iN} \right\} \quad (24)$$

$$\bar{h}_{\tilde{\lambda}}(\wp^*, O) = \bar{\phi}_N(T) = \max_{2 \leq i \leq N-1} \left\{ \bar{\phi}_i(T) * \bar{a}_{iN} \right\} \quad (25)$$

if the center of the interval as the defuzzified values is used, then

$$\wp^* = \arg \max_{\wp} \left(\frac{\underline{h}_{\tilde{\lambda}}(\wp, O) + \bar{h}_{\tilde{\lambda}}(\wp, O)}{2} \right) \quad (26)$$

similarly $\tilde{\phi}_j(t)$ and $h_{\tilde{\lambda}}(\varphi^*, O)$ 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 .

Finally, the IT2 fuzzy forward-backward algorithm (Baum-Welch) is utilized to refine the parameters of the initialized IT2 FHMM. The steps to perform parameters re-estimation are summarized as follows:

$$\hat{\mu}_{jm} = c \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) o_t^r}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)} \quad (27)$$

$$\hat{\sigma}_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) (o_t^r - \hat{\mu}_{jm})(o_t^r - \hat{\mu}_{jm})'}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)} \quad (28)$$

$$\hat{c}_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_j^r(t)} \quad (29)$$

$$L_{jm}^r(t) = \frac{1}{\xi(h_{\tilde{\lambda}}(O^r))} \xi \left(\tilde{a}_{jm}(t) \Pi \tilde{\beta}_{jm}(t) \right) \quad (30)$$

$L_{jm}^r(t)$ 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. After that, calculate the IT2 fuzzy forward and backward variables of the IT2 non-singleton fuzzified observation sequence O^r , $1 \leq r \leq R$, 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

$$\Lambda = \left(\sum_{r=1}^R (h_{\tilde{\lambda}}(\varphi^*, O^r) + \bar{h}_{\tilde{\lambda}}(\varphi^*, O^r)) \right) / 2 / R \quad (31)$$

for this iteration is not higher than the value at the previous iteration then stop, otherwise repeat steps using the new IT2FHMM (R represents the total number of observations).

(2) *Recognition Phase*

The main job of the classifier or recognizer is to classify which trained class the presently obtainable testing gesture belongs to. The recognizer has to mark those trained classes according to the presented testing gesture and the maximum matching score is the corresponding trained gesture class [4]. After finishing the training process by calculating the T2FHMM parameters for each type of gesture, a given gesture is recognized corresponding to the greatest likelihood of HMM model by using the Viterbi algorithm. In this case, the hand gesture is recognized by choosing 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 the Viterbi algorithm. The



FIGURE 2. Examples of sign images used in IT2FHMM model training

image is recognized as the hand gesture, whose model has the highest production probability [5].

To conclude, the system has presented the integration of type-2 fuzzy logic to the Hidden Markov Models. This integration consists of replacing the basic arithmetic operations by the adequate type 2 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 additive constraint, necessary for the HMM, towards the monotonicity one, much less restrictive.

4- Experimental Results

The accuracy and performance of the proposed IT2FHMM sign language recognizer are further verified using an experimental image dataset consisting of large vocabulary with more than 6000 signs taken from many sign language dictionaries with 10 samples per sign. The utilized database was downloaded from the standard benchmark datasets such as static hand posture database II with complex backgrounds (<http://www.idiap.ch/resource/gestures/>), hand image dataset (ASL rendered) (<http://www.cs.bu.edu/groups/ivc/data.php>), and hand action and gesture database (<http://homepages.inf.ed.ac.uk>). Figure 2 shows samples of hand images. Those samples are subjected to illumination changes, scaling, blurring, rotation, view-point variant and translated in uncontrolled background. 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. Five randomly selected samples per signs are used as the training samples while two of the remaining five samples are referred to as the registered test set (*Reg*), and the rest of the samples are referred to as the unregistered test set (*Unreg*). The recognition rate is calculated as the percentage of the number of correctly classified hand postures to the total number of hand postures [31]. 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.

Model	Recognition Rate (%)		
	HMM	FHMM	T2FHMM
6 states /10 mixtures	76.1	88.1	95.7
6 states / 3 mixtures	67.7	72.7	85.6
4 states /2 mixtures	59.5	70.5	82.3
6 states /6 mixtures	73.3	75.3	87.7
6 states /4 mixtures	69.4	79.4	91.9
Overall average	69.2	77.2	88.7

TABLE 1. Comparison results of different HMM models with number of states and mixture for unregistered test set.

Table 1 shows the test results of HMM, FHMM, and IT2FHMM-based isolated SLR performances on large vocabulary with different a number of states and different number of Gaussian mixture per state under the same observation vector of 52 SVD elements per sign. For IT2FHMM, the uncertainty factor $k=3$ and the range of transition $[\underline{a}_{ij}, \bar{a}_{ij}] = [0.98 \times a_{ij}, 1.02 \times a_{ij}]$. The average recognition rates of 95.5% for IT2FHMM, 82.3% for FHMM and 74.1% for HMM are observed for the registered test set. For the unregistered test set, the average recognition rates of 88.7%, 77.2%, 69.2 are obtained respectively. Experiments show that IT2FHMM increases the recognition accuracy by 10%, 20% than FHMM and HMM on the registered and unregistered (un-embedded training) test sets with large vocabulary size and large variations in sign data. The possible reason is that the combination of higher uncertainty ability of interval Type-2 fuzzy sets and excellent temporal processing properties of HMM in a novel scheme may compensate each other to obtain better results. In other words, the proposed system offers a promising potential to solve difficult singer independent hand gestures recognition in the case of large vocabulary size.

Since there is no algorithm to catch the ideal number of hidden states and mixture components, in order to attain as good a performance as possible (depending on the complexity of the gesture shape). The second experiment studies how the system performances change while varying the number of states and mixture components. The results are presented in Table 1. As Table 1 shows, the optimal number of states and mixture components are six and ten respectively. This experiment is conducted with 5 samples for each gesture class as a training set. Generally, these numbers are changed in diverse datasets because they are directly affected by observation distribution. Furthermore, results in Table 2 demonstrate 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. These optimal IT2FHMM parameters are of course valid strictly for the used database. If another database is used, these parameters will change within certain limits.

The third experiment compares the proposed system with that done in [33] for signer independent Arabic sign language. In spite of the fact that they use

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

TABLE 2. Recognition rate of different size of SVD feature for registered test set.

different feature extraction methods, setup, and classifier, both systems use the same datasets. As shown in Table 3, the suggested system performs much better than the DCT coefficient-based system (referred to as the ArSL– DCT coefficient-based system in Table 3). This is due to the fact that T2FHMM can deal with modeling parameters that may not accurately reflect the underlying distribution of the observations because of insufficient or noisy data in sign recognition problems.

Experiment four investigates how recognition rate of the proposed system depends on T2FHMM topologies [36]: Fully Connected (Ergodic model) where any state in it can arrive at any other states, Left-Right (LR) model such that each state can go back to itself or to the next states and Left-Right Banded (LRB) model that also each state can go back to itself or the following state only. From Table 4, the average ratio of LRB topology for 6 states was 88.7%. LRB topology was better than LR and Ergodic topologies. This result can be explained on the basis of that LRB is good for modeling-order-constrained time-series whose properties sequentially change over time. Since the model has no backward path, the state index either increases or stays unchanged as time increases. Since each state in Ergodic topology has many transitions than LR and LRB topologies, the structure data can be lost easily. In addition, LRB topology is more restricted rather than LR topology and simple for training data that will be able to match the data to the model and the number of states is decided on the basis of complexity of a gesture.

In experiment five, the accuracy of obtained IT2FHMM recognizer with optimal parameters that have been obtained experimentally can be assessed by a confusion matrix, opposing assigned class (column) by the classifier with their true original class (row) as shown in Table 5 of randomly selected samples. Entries along with the main diagonal are correct recognitions. Entries other than those on the main diagonal are recognition errors. The obtained test results exhibit accurate sign language recognition with low false alarms; thus, showing the robustness of the proposed system. It is noticeable that some signs result in particularly low recognition rates. The gesture “B”, for example, has a recognition rate of 70% and is mostly classified as “C”. This is due to the fact that the location, movement, and orientation of the central hand are very similar in both gestures. Therefore, the observation (feature) vectors formed from the hand tracking segment are most to be expected very close to each other. Thus, the system will become disordered between these two signs and provide relatively higher error rate for these specific gestures.

	Instruments used	Feature vector length	Average Recognition rate
Proposed system (SVD coefficients + T2FHMM Classifier)	None: Free hands	20 features only from SVD vector	92.4%
ArSL System (HMM Classifier +Feature-based hand Model)	None: Free hands	8 features per sign	83.2%

TABLE 3. Comparison with similar free hands signer independent isolated SLR system.

IT2FHMM	Topology		
	Ergodic	LR	LRB
Average Recognition Rate	69.48	85.45	88.7

TABLE 4. Isolated gesture recognition results from T2FHMM topologies with 6 states /4 mixtures.

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

TABLE 5. Confusion matrix (10 samples per class) .

4. Conclusions

This paper presents a novel approach based on higher order fuzzy HMM to increase recognition accuracy of signer independent isolated SLR. T2FHMM-based framework is used for robust estimation of the individual recognition units from the obtained SVD feature sequence. The set of images is trained by the forward-backward algorithm then we create the database for matching the input image to the system by the 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, easy to use and it works in real-time. It is fairly robust to the difficult background. The proposed method's results show the potential of the system to correctly recognize a large variety of the signs no matter what is the orientation of the hand in a particular sign. The future orientation concerns the

use of other hand feature extraction techniques to increase recognition rates and deal with continuous SLR.

REFERENCES

- [1] M. AL-Rousan, K. Assaleh and A. Talaa, *Video-based signer-independent arabic sign language recognition using hidden markov models*, Applied Soft Computing, **9**(2009), 990-999.
- [2] V. Athitsos and S. Sclaroff, *Estimating 3d hand pose from a cluttered image*, Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, USA, (2003), 1-8.
- [3] S. Bilal, R. Akmeiliawati, A. Shafie and M. Salami, *Hidden markov model for human to computer interaction: a study on human hand gesture recognition*, Artificial Intelligence Review, **40**(4)(2013), 495-516.
- [4] N. Binh, E. Shuichi and T. Ejima, *Real-time hand tracking and gesture recognition system*, Proc. Int. Conf. Graphics, Vision and Image Processing, Egypt, (2005), 362-368.
- [5] X.-Q. Cao, J. Zeng and H. Yan, *Modeling uncertain speech sequences using type-2 fuzzy hidden markov models*, Lecture Notes on Computer Science, **4810**(2007), 315-324.
- [6] E. Celik, M. Gul, N. Aydin, A. Gumus and A. F Guneri, *A comprehensive review of multi criteria decision making approaches based on interval type-2 fuzzy sets*, Knowledge-Based Systems, **85**(2015), 329-341.
- [7] V. Christian and D. Metaxas, *Parallel hidden markov models for american sign language recognition*, Proc. IEEE Int. Conf. Computer Vision, Greece, **1**(1999), 116-122.
- [8] P. Doliotis, V. Athitsos, D. Kosmopoulos and S. Perantonis, *Hand shape and 3d pose estimation using depth data from a single cluttered frame*, Lecture Notes in Computer Science, **7431**(2012), 148-158.
- [9] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, *A Hidden markov model-based isolated and meaningful hand gesture recognition*, Int. J. Electrical, Computer, and Systems Engineering, **3**(3)(2009), 156-163.
- [10] W. Gao, G. Fang, D. Zhao and Y. Chen, *A chinese sign language recognition system based on softm/srn/hmm*, Pattern Recognition, **37**(12)(2004), 2389-2402.
- [11] W. Gao, J. Ma and J. Wu, *Sign language recognition based on hmm/ann/dr*, Int. J. Pattern Recognition and Artificial Intelligencet, **4**(5)(2000), 587-602.
- [12] A. Ghotkar and G. Kharate, *Hand segmentation techniques to hand gesture recognition for natural human computer interaction*, Int. J. Human Computer Interaction, **3**(1)(2012), 15-25.
- [13] N. Ibraheem, M. Hasan, R. Khan and P. Mishra, *Understanding color models: a review*, Journal of Science and Technology, **2**(3)(2012), 265-275.
- [14] Z. Jia and Z.-Q. Liu, *Type-2 fuzzy sets for pattern recognition: the state-of-the-art*, Journal of Uncertain Systems, **1**(3)(2007), 163-177.
- [15] A. Jmaa, W. Mahdi, Y. Jemaa and A. Hmadou, *A new approach for digit recognition based on hand gesture analysis*, Int. J. Computer Science and Information Security, **2**(1)(2009), 1-8.
- [16] M. Kelarestaghi, M. Slimane and N. Vincent, *Introduction of fuzzy logic in the hidden markov models*, Proc. Int. Conf. Fuzzy Logic and Technology, UK, (2001), 14-16.
- [17] W. Kong and S. Ranganath, *Towards subject independent continuous sign language recognition: a segment and merge approach*, Pattern Recognition, **47**(3)(2014), 1294-1308.
- [18] K. Kumar, M. Ramakrishna, and B. Prasad, *Feature extraction using sparse svd for biometric fusion in multimodal authentication*, Int. J. Network Security and Its Applications, **5**(4)(2013), 83-94.
- [19] S. Lung, *Multi-resolution form of svd for text-independent speaker recognition*, Pattern recognition, **35**(7)(2002), 1637-1639.
- [20] S. Mitra and T. Acharya, *Gesture recognition: a survey*, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, **37**(3)(2007), 311-324.
- [21] G. Murthy and R. Jadon, *A review of vision based hand gestures recognition*, Int. J. Information Technology and Knowledge Management, **2**(2)(2009), 405-410.

- [22] S. Ong and S. Ranganath, *Automatic sign language analysis: a survey and the future beyond lexical meaning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **27(6)**(2005), 873-891.
- [23] K. Patwardhan and S. Roy, *Dynamic hand gesture recognition using predictive eigen tracker*, Proc. Int. Conf. Computer Vision, Graphics and Image Processing, India, (2004), 1-6.
- [24] J. Rekha, J. Bhattacharya and S. Majumder, *Hand gesture recognition for sign language: a new hybrid approach*, Proc. Int. Conf. Image Processing, Computer Vision and Pattern Recognition, USA, (2011), 1-7.
- [25] R. Shrivastava, *A hidden markov model based dynamic hand gesture recognition system using opencv*, Proc. IEEE Int. Conf. Advance Computing, USA, (2013), 947-950.
- [26] H. Sola, J. Fernandez, H. Hagraas, F. Herrera, M. Pagola and E. Barrenechea, *Interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: toward a wider view on their relationship*, IEEE Transactions on Fuzzy Systems, **23(5)**(2015), 1876-1882.
- [27] M. Su, *Fuzzy rule-based approach to spatio-temporal hand gesture recognition*, IEEE Transactions on Systems, Man and Cybernetics, **30(2)**(2000), 276-281.
- [28] N. Tanibata, N. Shimada and Y. Shirai, *Extraction of hand Features for Recognition of sign language words*, Proc. Int. Conf. Vision Interface, Canada, (2002), 391-398.
- [29] V. Tataru, R. Vieriu and L. Goras, *On hand gestures recognition using hidden markov models*, Electronics and Telecommunications, **51(3)**(2010), 29-32.
- [30] V. Vezhnevets, S. Vassili and A. Andreeva, *A survey on pixel-based skin color detection techniques*, Proc. Int. Conf. Graphicon, Russia, **3**(2003), 85-92.
- [31] C. Vogler and D. Metaxas, *A framework for recognizing the simultaneous aspects of american sign language*, Computer Vision and Image Understanding, **81**(2001), 358-384.
- [32] H. Wang, M. Leu and C. Oz, *American sign language recognition using multi-dimensional hidden markov models*, Journal of Information Science and Engineering, **22(5)**(2006), 1109-1123.
- [33] A. Youssif, A. Aboutabl and H. Ali, *Arabic sign language (arsl) recognition system using hmm*, Int. J. Advanced Computer System and Applications, **2(11)**(2011), 45-51.
- [34] J. Zeng and Z. Liu, *Type-2 fuzzy hidden markov models*, Studies in Computational Intelligence, **591**(2015), 57-83.
- [35] J. Zeng and Z. Liu, *Type-2 fuzzy hidden markov models and their application to speech recognition*, IEEE Transaction on Fuzzy Systems, **14(3)**(2006), 454-567.
- [36] J. Zeng, L. Xie and Z.-Q. Liu, *Type-2 fuzzy gaussian mixture models*, Pattern Recognition, **41**(2008), 3636-3643.

SAAD M. DARWISH 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 address: Saad.darwish@alex-igsr.edu.eg