Recognition System for Home-Service-Related Sign Language Using Entropy-Based *K*-Means Algorithm and ABC-Based HMM

Tzuu-Hseng S. Li, *Member, IEEE*, Min-Chi Kao, and Ping-Huan Kuo

***Abstract*—This paper presents a recognition system for understanding the words of home-service-related sign language. Because the data received from a sensor are sequential, the hidden Markov model (HMM) that has been successfully applied to speech signals is chosen as a classifier. However, the number of states in the HMM model should be decided upon first before constructing the HMM classifier. To solve this problem, an entropy-based *K*-means algorithm is proposed to evaluate the number of states in the HMM model with an entropy diagram. Four real datasets are utilized to verify the developed entropy-based *K*-means algorithm. Moreover, a data-driven method is given to combine the artificial bee colony algorithm with the Baum–Welch algorithm to determine the structure of HMM. The database contains 11 home-service-related Taiwan sign language words and each word is performed ten times, five males and five females are invited to perform such words. Finally, the recognition system is established by 11 HMM models, and the cross-validation demonstrates an average recognition rate of 91.3%.**

***Index Terms*—Artificial bee colony (ABC) algorithm, entropy-based *K*-means algorithm, hidden Markov models (HMMs), Taiwan sign language (TSL).**

I. INTRODUCTION

OVER the past decades, a considerable number of studies have been conducted on the interaction between humans and robots [1]–[5] or the environments [6]–[9]. However, this communication framework has been limited to a normal human who can speak to the robot directly. But there are some hearing-impaired people who cannot order the robot, using the spoken word, to serve them. This issue has encouraged us to develop a communication system for the hearing impaired so that the home-service robot can also assist them in their daily lives at home.

Deaf people use deaf sign language to communicate with each other. As with spoken language, each country has its own deaf sign language. There have been many studies on the recognition of deaf sign language, for example, Taiwan sign language (TSL) [10], American sign language (ASL) [11], British sign language [12], Italian sign language [13], Chinese sign language (CSL) [14], Korean sign language [15], and Arabic sign language [16].

Many researchers have proposed an application for sign language recognition. Chiu *et al.* [17] proposed a method to translate CSL to TSL and then synthesize sign videos. The grammar of sign language for deaf people is different from that of the spoken language or written language for hearing people. Wu *et al.* [18] proposed a Predictive Sentence Template (PST)-based language model to translate sign language into written language. These applications help hearing people to communicate with deaf people. Another application is translating sign language into sound. Hernandez-Rebollar *et al.* [19] applied hidden Markov models (HMMs) and neural network to translate isolated gestures of ASL into spoken and written words. Liang and Ouhyoung [20] developed a gesture recognition system to help the hearing impaired as well as the speaking impaired. Their system can translate ASL to speech. Gao *et al.* [21] constructed an interesting system, HandTalker, which can help deaf people make connections between gesture language and spoken language.

There are many sensors that can be used to acquire sign language data. According to the characteristics of the sensor, the collected data type is also different. The most frequently used sensor is the vision-based sensor, which can capture an image and analyze the movement trajectory of hands and facial expressions. Habili *et al.* [22] proposed a universal color-model and used motion cues to segment the face and hands from the sign language video sequence. Reference [23] extracted the videos and audio features to recognize human actions in realistic scenarios. Reference [24] extended vision-based solutions to recognize continuous signing. Vision-based approaches are natural and cost less when compared with glove-based approaches, which are the most expensive ones. However, vision-based approaches require great effort when removing the influences from the environment, these algorithms are relatively more complicated, and the computational load is also large.

Data gloves can acquire the most useful and precise data. Mehdi and Khan [25] first applied data gloves to sign language and translated ASL into an English sentence. Fang *et al.* [26] did excellent work; they solved the large vocabulary continuous sign problems by using data gloves. These gloves can acquire information on the structure of each finger and the position and

Manuscript received May 19, 2014; revised November 21, 2014; accepted March 14, 2015. This work was supported in part by the Ministry of Education, Taiwan, through the Aim for the Top University Project to the National Cheng Kung University, and in part by the Ministry of Science and Technology under Grant MOST 101-2221-E-006-193-MY3. This paper was recommended by Associate Editor S. Nahavandi.

The authors are with the aiRobots Laboratory, Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: thsli@mail.ncku.edu.tw; n2696421@gmail.com; coll22000@hotmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2015.2435702

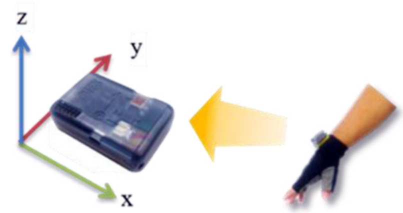

Fig. 1. AHRS sensor.

orientation of both hands. Although data gloves can acquire a lot of useful information, the cost is high.

The latest novel sensor is electromyogram (EMG), which can capture the activation signals of a muscle. Kosmidou *et al.* [27] proposed a method to extract 16 features from the EMG signals and evaluate their efficiency in ASL. Another often used sensor is micro-electro mechanical system (MEMS) accelerometers (ACCs). Bui and Nguyen [28] developed a device based on ACCs, which can recognize Vietnamese sign language. Some researchers combined the ACCs and EMG sensor to acquire more useful information at a lower cost when compared with that of data gloves. Kosmidou and Hadjileontiadis [29] acquired data from five-channel surface EMG and 3-D ACC from the signer's dominant hand to recognize isolated signs of Greek sign language. The mean classification accuracy was more than 93%. Zhang *et al.* [30] demonstrated the complementary functionality of ACC and EMG sensors and then detected the intensity of the EMG signals to detect the start points and end points of the meaningful gestures.

Another special sensor, the x-io sensor, can be run in Inertial Measurement Unit (IMU) mode or Attitude and Heading Reference System (AHRS) mode. In the IMU mode, the yaw angle will drift over time because the sensor only utilizes the gyroscope (Gyro) and ACC to measure the angular velocity and acceleration in each axis, and these values cannot compensate for the drift of the yaw angle. However, in the AHRS mode, the sensor utilized the magnetometer to calibrate the yaw angle. In this paper, we adopt x-io, which runs in AHRS mode, to acquire more accurate angular velocity, acceleration, and Euler angles in each axis when the hand moves. The Euler angle information helps us to know the orientation of the hands. Utilizing the AHRS sensor has the advantage that people do not need to stand in front of the vision-based recognition system and can call the home-service robot even if the robot is not in the sample place with the people. The equipment for the experiment is shown in Fig. 1.

Many approaches are submitted to recognize sequential data. Dynamic time warping (DTW) and longest common subsequence (LCS) are two methods, which can serve as the metric for similarity of sequential data. Lichtenauer *et al.* [31] combined the statistical DTW and quadratic classification on Discriminative Features Fisher mapping classifier to recognize Dutch sign language, with an average accuracy of 92.3%. Reference [32] utilized the LCS approach to recognize user-defined gestures, and humans can communicate with the robots via gestures. However, DTW-based and LCS-based methods are only suited for small datasets since they apply template-matching techniques to measure the similarities between the input data and the sequences in the dataset. If the number of sequences in the dataset is very large, the computation will take a lot of time.

Comparing DTW and LCS, another approach is to construct a classification model. Reference [33] indicated a hierarchical change detection test, which coupled with K Nearest Neighbor (KNN) and Support Vector Machine (SVM), is an appropriate tool for processing and classifying sequential data. Different from DTW and LCS that do not need to construct a model, the classification model spends a lot of time training a classifier in the training phase, but the recognition result can be obtained quickly in the testing phase. Vamplew [34] developed a sign language recognition system using neural network. However, the most popular classifier to recognize sequential data is HMM [35]–[37], which can model the data variation with time. Fang *et al.* [38] merged the decision trees and a self-organizing feature maps/HMM (SOFM/HMM) classifier to recognize CSL. SOFM, which serves as an implicit different signers' feature extractor for continuous HMM, is proposed as a special component of a fuzzy decision tree to get the final results at the last leaf nodes. In recent years, many researchers have also proposed certain kinds of knowledge-based HMM structures in many applications. Bouchaffra [39] proposed conformation-based HMMs, which can obtain the shape information from visible observation sequence to identify a human face. Antwarg *et al.* [40] built a tree whose node has the same structure as HMM but with different transition probabilities. The researcher called this algorithm an attribute-driven HMM tree. The algorithm is used to discover the relevant attributes and is tested by two real datasets, a Web application, and a mobile application dataset. Wu *et al.* [41] developed a two-level hierarchical alignment-based semicoupled-HMM to determine an optimal emotional state and this model can be used to construct an audiovisual emotion recognition system. Li *et al.* [42] applied particle swarm optimization to learn the HMM parameters and tested this classifier through the University of Central Florida human action dataset [43], which consists of 60 instances of common activities performed in an office environment including seven classes: 1) open door (18 instances); 2) pick up (21 instances); 3) put down (17 instances); 4) close door (four instances); 5) erase board (four instances); 6) pour water into cup (three instances); and 7) pick up object and put down elsewhere (eight instances).

In this paper, we apply the entropy index to evaluate the number of states in each HMM and then learn the structure of the HMM using the artificial bee colony (ABC) algorithm. The overall recognition system is shown in Fig. 2, where the classifier is tested by a home-service dataset, which consists of 11 TSL words with 1100 data. The major contributions of this paper are as follows.

1) Constructing a recognition system that can realize 11 home-service-related words of TSL.
2) Proposing a visualization method with entropy diagram, which can be served as a reference to evaluate the number of clusters in the dataset.
3) Presenting a novel method that combines the Baum–Welch algorithm with ABC algorithm to determine the structure and train the parameters of HMM.

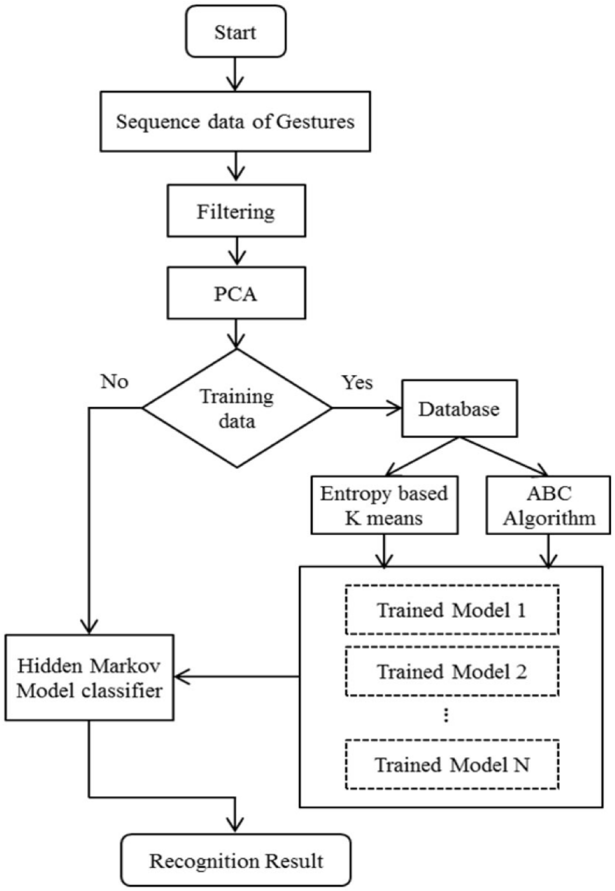

Fig. 2. System diagram of TSL Recognition.

This paper is organized as follows. In Section II, the preprocessing filters, principal component analysis (PCA), and entropy-based *K*-means algorithm of the sequence data are introduced. The model construction of recognition system for home-service-related TSL is described in Section III. Daily home-service sign language words used in this paper are shown in Section IV. In Section V, the experiments with entropy-based *K*-means algorithm and HMM classifier are presented to illustrate the effectiveness of the designed recognition system. Section VI is the conclusion of this paper.

II. Preprocessing of Sequence Data

Before constructing the recognition model, it is worthwhile and necessary to perform a prior process for the sequence data so that one can acquire more clear, reliable, and useful features for the classifier. The preprocessing procedure is examined in the following section.

A. Filtering

The receiving data from the AHRS sensor is not clear and has little variation even when the sensor is stationary. To get more accurate and useful data, low-pass filter (LPF) and Kalman filter are utilized in this paper.

The acceleration received from the AHRS sensor is polluted by noise. To remove the high-frequency noise, we compare the result of LPF and Kalman filter to decide which one is better to adopt. As shown in Fig. 3, the LPF brings out cleaner results in the acceleration value than does the Kalman filter.

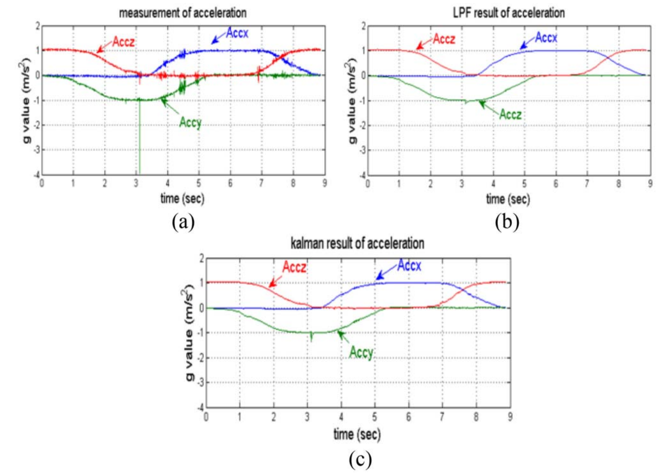

Fig. 3. (a) Raw acceleration data of the sensor. (b) LPF result of acceleration data. (c) Kalman filter result of acceleration data, which is measured in *z*-*y*-*x*-*z*-axis order. The impulse in third second is caused by the sensor hitting the desk.

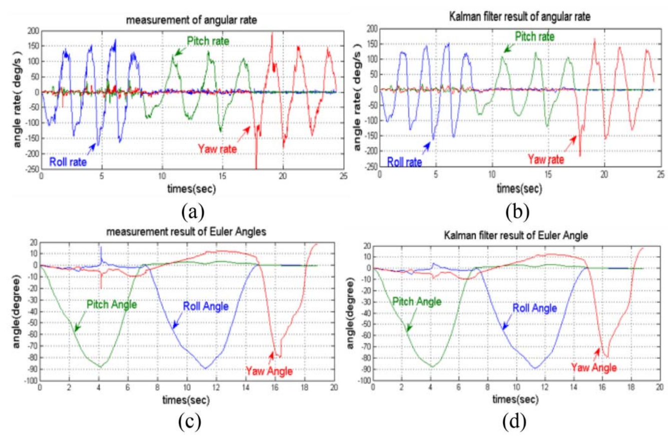

Fig. 4. (a) Raw Gyro data of the sensor. (b) Kalman filter result of the Gyro data acquired by rotating the sensor between 0° and −90° in three axes by hand. (c) Raw Euler angle data. (d) Kalman filter result of Euler angle.

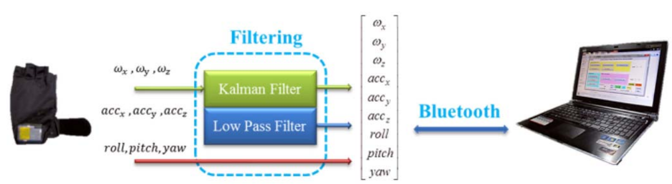

Fig. 5. Filtering of IMU data.

Moreover, the received Gyro data fluctuates wildly as shown in Fig. 4(a). To remove the noise and remain a dynamic property, this paper uses Kalman filter to estimate the real value of the angular velocity. Kalman filter can calibrate the prediction with the prediction error of the measurement and get the final estimate. Tuning the parameters related to the Kalman gain, one can adjust the proportion of the predicted value to the measurement. Fig. 4(b) shows the Kalman filter result of Gyro value. Fig. 4(c) and (d) illustrates the raw Euler angle data and Kalman filter result, respectively. From the experiment results, the Euler angles received from the AHRS sensor are accurate enough, so we do not have to apply any filter to the Euler angles. Fig. 5 shows the filters used in this paper.

B. PCA

The goal of PCA is to cancel the redundant dimension and minimize the number of dimensions of feature spaces without losing the main information so that we can represent each data with simplicity and decrease the computation loading. Each dimension found by PCA is independent and unique, which means that no correlation between each dimension and each component has a distinct meaning. Reducing the dimension of the feature space can avoid the curse of dimensionality. Thus, one does not need a large number of training data in training phase and this can speed up the training time and meet the requirement of real time. PCA is also a good method to find out the representation of the signals in a lower dimension. Therefore, we transform the received data from the AHRS sensor to another feature space via PCA and let the data representation be more efficient.

The suitable dimensions should be also determined by the succeeding methods, entropy-based K-means algorithm, and ABC-based HMM. These three procedures, PCA, entropy-based K-means algorithm, and ABC-based HMM should operate in coordination to decide the suitable dimensions and achieve satisfactory recognition results.

C. Entropy-Based K-Means Algorithm (Automatically Deciding on the K)

Before constructing the HMM model in the later sections, determining how many states are suitable for the model is important. The cluster number is related to whether the model is coarse or fine. More clusters mean more parameters can be tuned and the model behavior is more flexible and reliable. However, it also means numerous data are needed to acquire and adjust those parameters. Conversely, if the cluster number is too small, the model behaves stiffly and cannot reach satisfactory performance. For the reason mentioned above, it is evident that deciding on a suitable number of clusters can be helpful in constructing an acceptable model.

Malyszko and Stepaniuk [44] proposed the rough entropy-based K-means algorithm, which has been applied to the area of image segmentation. The rough entropy is used to determine the parameters, weights, in the rough K-means algorithm. Finally, the researchers selected three color images from Berkeley image dataset to verify the algorithm. Bai *et al.* [45] proposed the entropy-based soft K-means algorithm clustering method that utilizes the entropy and relative entropy information to guide the training process. Researchers used the entropy quantity to control the stiffness parameter β and get a more clear data partition. The relative entropy quantity is used to find a suitable number of clusters. However, Bai *et al.* [45] used only one dataset to validate the algorithm and listed entropy and relative entropy matrix, respectively, for various β values in two different cases ($K = 3$ and $K = 4$).

The algorithms proposed in [44] and [45] are not intuitive enough for visualization. Therefore, in this paper, we propose a method to determine the cluster number automatically with entropy diagram and validate the algorithm with two artificial datasets and four real datasets.

1) Adjustment of the K-Means Result: In this paper, we propose a criterion to decide the number of clusters. However,

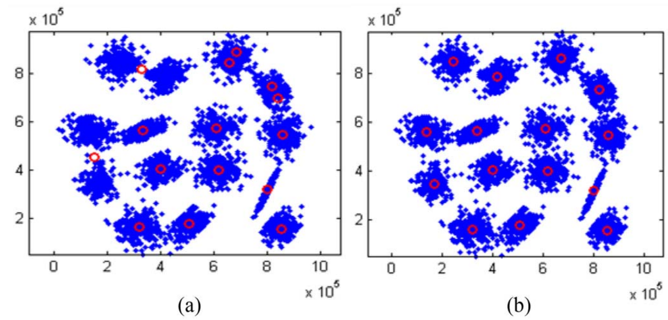

Fig. 6. (a) Clustering result of traditional K-means algorithm. (b) Clustering result of modified K-means algorithm.

before determining the cluster number, it is important to guarantee that the centers found by K-means algorithm locate in the central region of each cluster. If the centers found by K-means algorithm are not in the centroid of the cluster, but are in fact in the middle position of the two clusters, the proposed criterion may produce errors.

To ensure that the center calculated by K-means algorithm locates in the centroid of the cluster, this paper proposes two parameters to adjust the final clustering result. The first parameter is distance scale s_d, which means distance scale and can be set by the user manually, while the second parameter is sample scale s_s, which presents the proportion of samples in the restricted area to the cluster. Distance scale s_d and sample scale s_s should operate in coordination to make sure that the distribution of sample points inside each cluster is compact. Once the K-means algorithm has been performed, the clustering result should be examined by distance scale s_d and sample scale s_s. For example, suppose $s_d = 1/3$ and $s_s = 1/2$, which means that the cluster should include at least 1/2 sample points of all points inside the cluster around 1/3 circular range of distance. To avoid the center locating in the sparse region as shown in Fig. 6(a), we compute the distance threshold for each cluster k as

$$\mathrm{DTH}_k\big(x_i, x_j, c_k, s_d\big) = \left(\frac{(x_i - c_k)^2 + \big(x_j - c_k\big)^2}{2}\right) \times s_d \quad (1)$$

where DTH_k denotes the distance threshold of cluster k, x_i, and x_j, which are two farthermost points away from the center c_k, and s_d is a scalar between 0 and 1, which means distance scale

$$M_k = \{x_1, x_2, \ldots, x_m\} \quad (2)$$

where M_k is the set of points whose distance from center c_k is less than DTH_k. Suppose we have z clusters, and

$$\frac{|M_1|}{N_1} \le \frac{|M_2|}{N_2} \le \cdots \frac{|M_k|}{N_k} \cdots \le \frac{|M_z|}{N_z} \quad (3)$$

where N_k is the total number of samples in cluster k. Fig. 6(b) shows the clustering results of modified K-means algorithm and Fig. 7 illustrates the procedures of how to adjust the clustering result of K-means algorithm. If the smallest sample proportion of cluster 1 is less than the setting sample scale, execute K-means algorithm again until the smallest sample proportion of the cluster is also larger than the setting sample

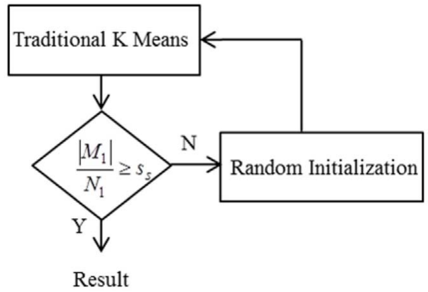

Fig. 7. Adjusted K-means algorithm.

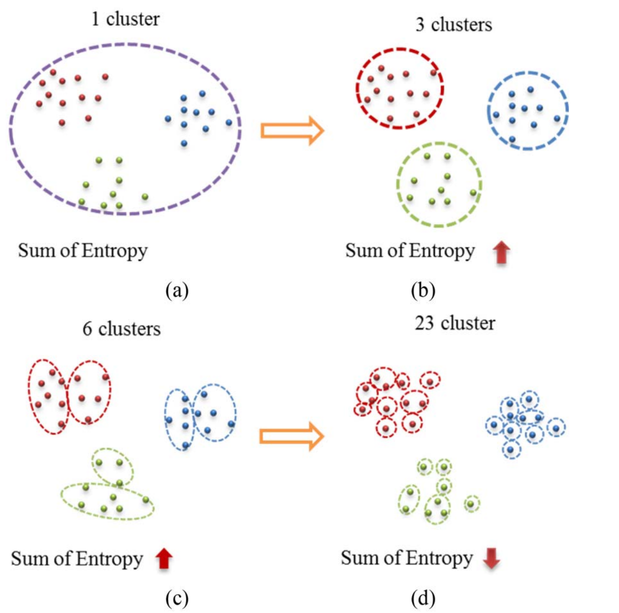

Fig. 8. Sum of entropy will vary with the number of clusters. (a) The entropy value is low because the data concentrate in certain regions. (b) The entropy value is high because the data are uniformly distributed. (c) The entropy value increases from (b) to (c) because one cluster is decomposed into two clusters. (d) The entropy value decreases when the number of clusters becomes larger and larger.

scale. Before performing the traditional K-means algorithm, the initial centers of the clusters will be randomly reselected.

2) Criterion for Evaluating K: This paper proposes a criterion, entropy index, to evaluate the value K of the K-means algorithm. The criterion can provide a reference for deciding the number of clusters. Setting the parameters s_d and s_s appropriately, we can find out the most possible cluster number by observing the entropy diagram. The entropy is defined by

$$\begin{aligned} \text{entropy} &= \sum_{k=1}^{z} H_k(P) = \sum_{k=1}^{z} \sum_{x_j \in k} -P(x_j) \log_2 P(x_j) \\ &= \sum_{k=1}^{z} \sum_{x_j \in k} \sum_{d=1}^{m} -P(x_{j_d}) \log_2 P(x_{j_d}) \end{aligned} \tag{4}$$

where $H_k(P)$ denotes the entropy of cluster k, P denotes the probability distribution, $P(x_j)$ denotes the probability of point x_j, and d denotes the dimension.

Fig. 8 is a diagram to explain the criterion. To use the entropy criterion, two points must be considered. The first is the calculation of the entropy. To calculate the entropy in (4), one needs to know the data distribution first. According

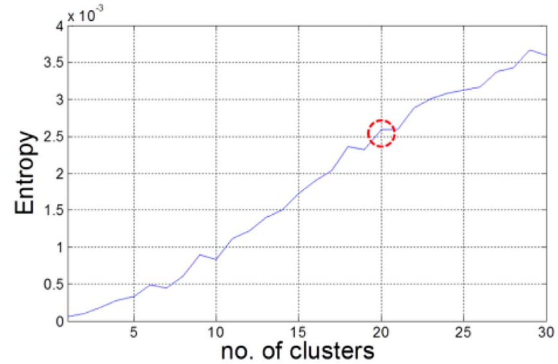

Fig. 9. A1 dataset (2-D, 3000 vectors, and 20 clusters) [47].

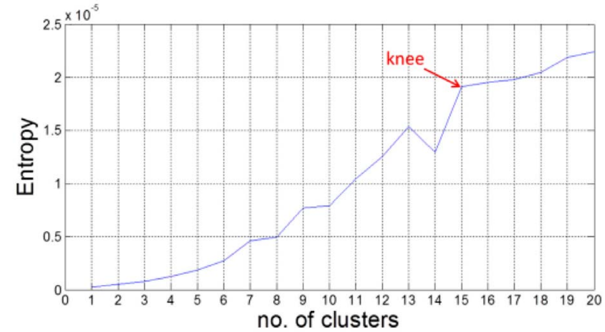

Fig. 10. S-sets (2-D, 15 Gaussian clusters, and 5000 vectors) [47].

to the maximum entropy criterion, if the data distribution is unknown, the possible choice is the uniform distribution [46]. This assumption may cause the estimation of the cluster number to deviate from the true number of clusters since the true data distribution is not uniform distribution. The second is to define the feature point of the entropy diagram. In general, the maximum point or the point whose slope varies tremendously is a good choice. Note that the diagram just provides a reference for evaluating the cluster number, but selecting which point serves as the cluster number is decided by the user.

Since the entropy is to measure the randomness of the data, we can conclude that if the data spreads uniformly in a cluster, as shown in Fig. 8(b), the entropy value is larger than the entropy in Fig. 8(a), which displays the data concentrated in the local region and not randomly enough as in Fig. 8(b). In Fig. 8(b), the data is distributed uniformly everywhere, which means that the degree of randomness is high. The entropy increases from Fig. 8(b) to (c). The reason for this is that one cluster is decomposed into two clusters. Figs. 9 and 10 are two illustrations of real datasets [47] that demonstrate this phenomenon. As Figs. 9 and 10 show, we discover that the entropy value increases with the number of clusters. Suppose the dataset is divided into more and more clusters, then the data distribution of each cluster is similar to uniform distribution. According to Jensen's inequality in information theory [48], when the entropy value calculated by Gaussian distribution grows, this implies that the real-data distribution might become uniform. However, the entropy value will fall when the number of clusters becomes larger and larger as shown in Fig. 8(d). In Fig. 8(d), each cluster contains only one to three samples and the number of samples in each cluster

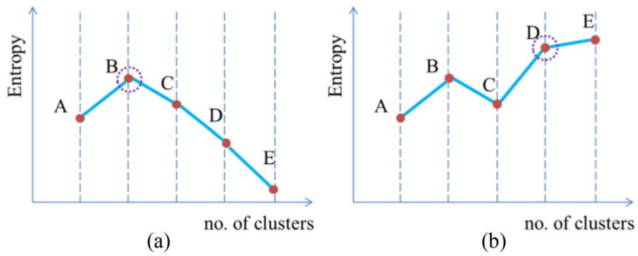

Fig. 11. Example of (a) maximum entropy criterion and (b) knee criterion.

is very low, so the probability of each point is higher and the degree of randomness is low.

This paper proposes two criteria for deciding the number of clusters. The first criterion is the maximum entropy criterion. In general, because the distribution of collected data is unknown, the prior selection is the maximum value of the entropy as shown in Fig. 11(a). However, if the entropy value trends toward increasing with the number of clusters, the maximum value of the entropy may not be a good choice. In this case, we suggest using the second criterion, the knee point of the entropy diagram. Fig. 11(b) shows that the entropy keeps increasing with the number of clusters, and two special points, B and D, should be noted. Point B is the maximum point from A to C. The fact that the entropy decreases from B to C means that a sparse region exists in the clustering result at point C; therefore, the clustering result of B is more uniform than that of C. Nevertheless, the entropy subsequently begins to rise from C to D again. From an observation of the segment from B to D, it can be seen that point D is considered more likely to be the real number of clusters because the clustering result of D is more uniform than that of B and C. From D to E, the increasing degree of entropy is very small, which implies that the entropy is trending to converge. Therefore, one can conclude that the clustering result at point D is more uniform than that at B and is more likely to be the real number of clusters. Point D is called the knee point. The following equation is proposed to search the knee point of the entropy diagram:

$$\text{num} = \arg\max \frac{[\text{entropy}(x) - \text{entropy}(x-1)]}{[\text{entropy}(x+1) - \text{entropy}(x)]} \tag{5}$$

where both $\text{sign}[\text{entropy}(x) - \text{entropy}(x-1)] = 1$ and $\text{sign}[\text{entropy}(x+1) - \text{entropy}(x)] = 1$.

III. Model Construction of Sign Language Recognition System

HMM [35] has been proven to be an excellent model to describe the sequential signal. In this paper, we use HMM to recognize the TSL words. In spoken language, the left-to-right structure of HMM is often used to model the speech signals according to the spelling rules. However, sign language cannot just simplify the HMM to left-to-right structure since sign language does not define the basic phoneme of a word and also does not have spelling rules. To solve the phoneme problem, we propose the entropy-based K-means algorithm and try to find out the phoneme-like unit for each sign word. Based on the entropy criterion, one can evaluate the number of clusters in each sign word data, and each cluster modeled by Gaussian distribution can be denoted as a phoneme. The next problem is to decide on the HMM structure. Many researches [35]–[37] have investigated various kinds of HMM structures with the knowledge-driven method to promote the recognition rate. In this paper, we propose a data-driven method, which combines the ABC algorithm and Baum–Welch algorithm to determine the structure of HMM.

A. Structure Learning of Data-Driven HMM

The sequential data can be represented by $\overline{O} = [\overline{o}_1, \ldots, \overline{o}_t, \ldots, \overline{o}_T]$, and each feature vector $\overline{o}_t$ is composed of D dimensions $\overline{o}_t = [x_1, x_2, \ldots, x_D]^T$. We use $\lambda = (N, A, B, \pi)$ to represent the HMM model. N is the total number of states and matrix $A = [a_{ij}]$ means the transition probability matrix, which can be defined by [35]

$$a_{ij} = \text{Prob}\big[q_t = j | q_{t-1} = i\big] \tag{6}$$

where q_t represents the state number at time t, so the a_{ij} means the transition probability from state i to state j. If the HMM model has N states, there are N rows in matrix B as described below [35]

$$B = \big[b_1(\overline{o}), b_2(\overline{o}), \ldots, b_j(\overline{o}), \ldots, b_N(\overline{o})\big]^T \tag{7}$$

$$b_j(\overline{o}) = \sum_{k=1}^{M} c_{jk} N\big(\overline{o}; \mu_{jk}, U_{jk}\big) \tag{8}$$

$$\sum_{k=1}^{M} c_{jk} = 1 \tag{9}$$

$$\pi = [\pi_1, \pi_2, \ldots, \pi_N] \tag{10}$$

where $b_j(\overline{o})$ is composed of M multivariable Gaussian distribution, c_{jk} is the weight for each Gaussian distribution, and μ_{jk} is the mean vector for the kth mixture component. U_{jk} is the covariance matrix for the kth mixture component. π means the initial probabilities of each state in this HMM. $P(\overline{O}|\lambda)$ is the probability of seeing the sequential data $\overline{O}$ given the HMM model λ. We use the conventional Baum–Welch algorithm [49] to adjust the parameter c_{jk}, μ_{jk}, and U_{jk}, because the algorithm can guarantee that $P(\overline{O}|\bar{\lambda}) \geq P(\overline{O}|\lambda)$ for each iteration. The parameters can be calculated by [49]

$$\gamma_t(j,k) = \left(\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}\right)\left(\frac{c_{jk}N\big(o_t; \mu_{jk}, U_{jk}\big)}{\sum_{m=1}^{M} c_{jm}N\big(o_t; \mu_{jm}, U_{jm}\big)}\right) \tag{11}$$

where

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T}\sum_{k=1}^{M} \gamma_t(j,k)} \tag{12}$$

$$\bar{\mu}_{jk} = \frac{\sum_{t=1}^{T}\big[\gamma_t(j,k) \bullet o_t\big]}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{13}$$

$$\bar{U}_{jk} = \frac{\sum_{t=1}^{T}\Big[\gamma_t(j,k)\big(o_t - \mu_{jk}\big)\big(o_t - \mu_{jk}\big)'\Big]}{\sum_{t=1}^{T} \gamma_t(j,k)}. \tag{14}$$

B. ABC Algorithm

The ABC [50] algorithm is a newly developed evolutionary algorithm, which can be used to address the optimization problem. This algorithm imitates bee swarm behavior and tries to find the optimal solution. Each bee position x_{ij} represents the candidate solution for the optimization problem, where $i = 1, 2, \ldots,$ PS represents the candidate solution index and $j = 1, 2, \ldots,$ Dim represents the parameter index that would be tuned in this optimization problem. For example, x_1 represents the first candidate solution and x_i means the ith candidate solution. ABC defines three kinds of bee, the employed bee, the onlooker bee, and the scout bee, to search the parameter space.

At the initialization stage, one should determine the number of bees, namely, the population size PS, the number of parameters, the upper bound and lower bound of the parameter value, and the limit count for the scout bees. After setting these parameters, one can calculate the fitness value of the initial bees and compute the selected probability of each bee according to the fitness value. The selected probability is calculated by [50]

$$p(x_i) = \frac{\mathrm{fit}(x_i)}{\sum_{i=1}^{\mathrm{PS}} \mathrm{fit}(x_i)} \tag{15}$$

where $\mathrm{fit}(x_i)$ is the fitness value of x_i and is defined as follows:

$$\mathrm{fit}(x_i) = \frac{1}{-f(x_i)} \tag{16}$$

where $f(x_i)$ is the objective function value and the object function is defined according to the optimization problem.

The next stage is the parameter tuning phase, which includes the employed bee phase, the onlooker bee phase, and the scout bee phase. Each bee represents the candidate solution of the problem, and this stage is to tune the value of each bee to enhance the fitness value of the bees. The main difference between the employed bee and the onlooker bee phases is that the onlooker bee possesses a probability condition to decide whether its value should be adjusted. If the new solution generated by the employed bee phase or onlooker bee phase is better than the old one, one replaces the old solution with the new one.

If the new solution generated by the two methods is worse than the old one, we hold the old one but add 1 to the counter of the scout bee. The new solution can be generated as follows [50]:

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) \tag{17}$$

where v_{ij} represents the new solution and $\phi_{ij} \in [-1, 1]$ is a random number and k must be different from i. The ith bee is tuned by kth bee that is around the ith bee. In the scout bee phase, if the counter of a certain bee exceeds the limit of the setting, the algorithm regenerates a new solution to enhance the diversity of the bees.

C. ABC Algorithm-Based HMM Structure

For each TSL word, we constructed an HMM to model this gesture. For each gesture, we collected training data to train the model. When the observation data $\overline{O}$ and the initial model λ are received, the objective function $P(\overline{O}|\lambda)$ is to maximize by adjusting the parameter of λ. In this paper, the ABC algorithm and the Baum–Welch algorithm [49] are integrated to resolve this optimization problem.

First, entropy-based K-means algorithm is used to decide the number of states for each sign language HMM model, and the clustering result can be served as the initial value of mean vector μ_{jk} and the covariance matrix U_{jk} of each state. For example, suppose the state number of the HMM model of hello is N, then the number of entries in the transition probability matrix A is $N \times N$. Therefore, the bee x_{ij} has $N \times N$ parameters. In the past, the Baum–Welch algorithm was often used to search the optimal value of the parameters of HMM. However, the result calculated by the Baum–Welch algorithm is just a local optimum. To find a better value of the parameters, we adopt the ABC algorithm to tune the transition probability matrix A, which decides the structure of HMM. Suppose the bee colony size is PS, we calculate the fitness value of each x_i, $i = 1 \ldots \mathrm{PS}$, and select the best solution x_{best}. $P(\overline{O}|\lambda)$ can be described as follows:

$$P\left(\overline{O}|\lambda\right) = \sum_{i=1}^{N} P\left(\overline{O}, q_t = i|\lambda\right) = \sum_{i=1}^{N} [\alpha_t(i)\beta_t(i)]. \tag{18}$$

To reduce the computation loading and prevent underflow error, the objective function is redefined as follows:

$$g(\lambda) = \log P\left(\overline{O}|\lambda\right). \tag{19}$$

The ABC algorithm is used to learn the structure of HMM, λ, and to adjust the transition probability matrix. The entries of A are saved in x_{ij}, $j = 1, 2, \ldots, N \times N$, and the fitness function can now be defined by

$$\mathrm{fit}(x_i) = \frac{1}{-g(\lambda(x_i))}. \tag{20}$$

The selected-probability for each bee is calculated by

$$p(x_i) = 0.9\frac{\mathrm{fit}(x_i)}{\mathrm{Max}[\mathrm{fit}(x_i)]} + 0.1. \tag{21}$$

The proposed ABC-based HMM is depicted by Algorithm 1. Fig. 12 shows the training process using ABC algorithm to tune the structure of HMM models.

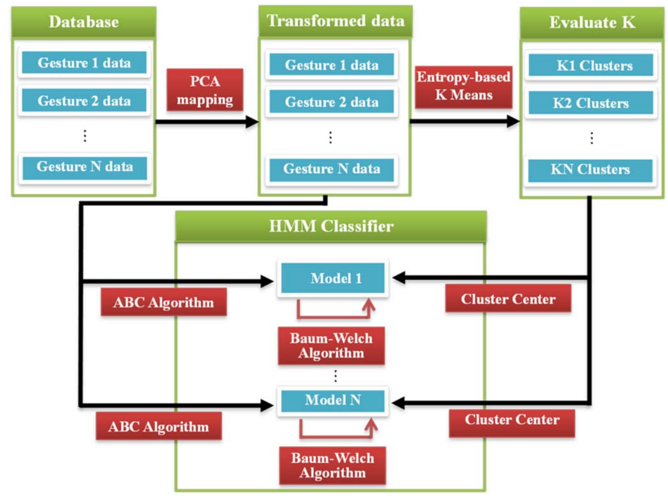

Fig. 12. Training process using the ABC algorithm to tune the structure of HMM models.

IV. Daily Home-Service TSL Words

The recognition system proposed in this paper includes 11 TSL words: hello, good morning, good afternoon,

Algorithm 1: ABC Algorithm Based HMM Structure
Step 0: Construct training set $\overline{O}$ for each HMM;
Step 1: Run Entropy-K-Means to get the state number N for each HMM, and the parameter number of objective function of this HMM is $N \times N$;
Step 2: Setting the Bee Colony Size, limit for scout, lower bound and upper bound of parameter, termination condition;
Step 3: Initialize ABC with parameters and objective function and generate food source $x_{ij} = randn(0 \sim 1), i = 1,\ldots, PS, j = 1,\ldots,N \times N$ **and then remember the Best food source** x_i **;**
Step 4: While the termination condition is not met, do Step 5~Step9;
Step 5: Send the EmployedBees **For i=1:PS** $x_{newj} \leftarrow$ do small perturbati on to x_{ij} (the small perturbati on can be considered as the bees around the food x_{ij}) $Normalizeeachrow\ of\ matrix\ A. In\ other\ words, \sum_{j=1}^{N} x_{newj} = 1$ **Calculate the fitness value** $fit(x_{newj})$ **IF** $fit(x_{newj}) > fit(x_{ij})$ **,** $x_{best} \leftarrow x_{newj}$ **&** $count(x_{ij}) \leftarrow 0$ **Else** $count(x_{ij}) \leftarrow count(x_{ij}) + 1$ **End**
Step 6: Calculate selected-probabilities of each Bee as equation (21) described;
Step 7: Send the OnlookerBees; $i = 0, t = 0$ **While** $t < PS$ $r \leftarrow randn(0 \sim 1)$ **IF** $r < \text{Prob}(x_{ij})$ $t \leftarrow t + 1$ **For i=1:PS** $x_{newj} \leftarrow$ perturbati on(x_{ij}) (the small perturbati on can be considered as the bees around the food x_{ij}) $Normalizeeachrow\ of\ matrix\ A. In\ other\ words, \sum_{j=1}^{N} x_{newj} = 1$ **Calculate the fitness value** $fit(x_{newj})$ **IF** $fit(x_{newj}) > fit(x_{ij})$ **,** $x_{best} \leftarrow x_{newj}$ **&** $count(x_{ij}) \leftarrow 0$ **Else** $count(x_{ij}) \leftarrow count(x_{ij}) + 1$ **End** **End** $i \leftarrow i + 1$ **End**
Step 8: Memorize Best food source x_{best} **;**
Step 9: Send ScoutBees ; **Determine the food source** x_i **whose trial counter exceeds the "limit" value.** $Max\_count = 0, x\_\text{limit} = 0$ **For i=1:PS** **IF** $counter(x(i)) > Max\_count$ **,** $Max\_count = counter(x(i))$ $x\_\text{limit} \leftarrow x(i)$ **End** **End** **IF** $Max\_count > \text{limit}$ $randn(x\_\text{limit})$ **End**
Step 10 return x_{best}
Step 11 Run Baum-Welch algorithm for each HMM
Step 12 Get the HMM Classifier which is composed of all HMM

good night, black tea, milk, coffee, tissue paper, living room, bedroom, and kitchen. Fig. 13(a) shows the actions of the greeting words, Fig. 13(b) depicts the words for daily necessities that people may use in their lives, and Fig. 13(c) illustrates the words representing the location in a home. Because the recognition system will be placed on a home-service robot, we hope that the recognition system can recognize the three kinds of words of TSL. The first type of word can be used to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

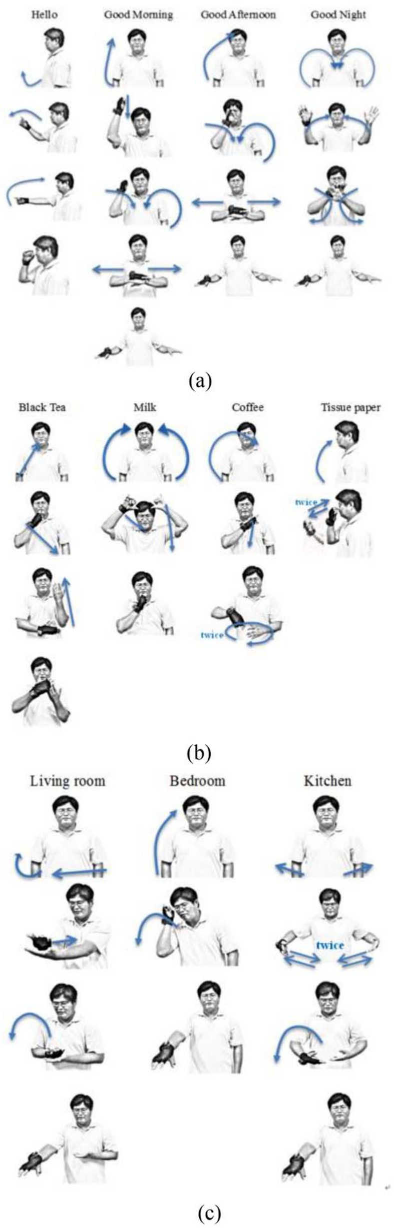

Fig. 13. Daily home-service TSL words. TSL for (a) greeting someone, (b) daily necessities, and (c) place.

greet someone, while the second type of word can be used to tell the robot what food or drink the human needs. The third type of word is the place in a home; a person can use this word to tell the robot where the person is.

V. System Setup and Experimental Results

This section describes how to set up the entire experimental system and how to collect the data and construct the database. Two experimental results with entropy-based K-means algorithm are given, one is for four real datasets, and the other is for sign language data. The last experiment illustrates the results of the proposed HMM structure.

A. Data Collection and Correction

Because the sensor data we used is not just ACC, pith, roll value, and Euler angles, it also includes yaw data, which is affected by magnetic force. This characteristic makes the sensor data untrustworthy if the person's standing orientation is different from that of the previous experiment. To address this problem, we must set a standard coordinate system to guarantee that the collected data is based on the same coordinate system every time. Because the sensor x-io [51], [52] we used has a built-in coordinate correction algorithm, which can be triggered by a user's preference, we needed to just decide on the starting position and when to perform this correction algorithm. For convenience, we chose the position of arms hanging naturally as the starting position, which makes the starting position at the side of the human body. The coordinate correction algorithm will be performed at the end of the action for each TSL word.

B. Construct the Database

In this paper, five men and five women are invited to perform the TSL word, each word ten times; and then the data were collected to construct a TSL database. The database consists of 11 home-service-related words which are shown in Fig. 13.

However, we know that the length of each sequential data received from the sensor will not remain the same, even though the same person performs the same TSL word. To judge when the gesture starts and ends, we calculated the energy of the input point by

$$\text{Energy} = \text{sqrt}\left(x_{\text{acc}}^2 + y_{\text{acc}}^2 + z_{\text{acc}}^2 + \omega_x^2 + \omega_y^2 + \omega_z^2\right). \qquad (22)$$

A threshold is set up for splitting the meaningful content of the TSL word from the received sequential data. That is, when the energy of the received data point was greater than that of the threshold, then the point would be put into the meaning-data sequence. The other threshold value is the length of the meaning-data sequence. The ending point is determined by the energy and length thresholds.

C. Verification of Entropy-Based K-Means Algorithm With Four Real Datasets

This paper applies four real datasets, breast-cancer [47], wine [47], glass [47], and yeast [47] datasets to demonstrate the effectiveness of the entropy-based K-means algorithm. It should be noted that this algorithm is not designed to achieve the goal of the correct recognition or correct clustering. However, what we really care about is finding the cluster number of the dataset. To apply the entropy-based method, one should decide which distribution is used to model the data. According to the maximum entropy criterion, we choose the uniform distribution as the data distribution when the distribution is unknown.

Table I lists the experimental results of the four datasets. Fig. 14 shows the diagram results of the entropy-based K-means algorithm. As mentioned previously, the peak of the entropy diagram often represents the number of clusters in this dataset. From the experimental results, we can find that the four datasets are suitable to be analyzed using the entropy-based method.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
ENTROPY-BASED *K*-MEANS ALGORITHM RESULT OF FOUR DATASETS

Distance_scale =1/3, Sample_scale =1/5, Uniform distribution					
Dataset	Normalize	PCA	Mean	Std.	Time
Breast-cancer (9 dimensions, 2 clusters)	Y	N	2.05	0.22	1.55s
Wine dataset (13 dimensions, 3 clusters)	Y	Y(1 component)	3.00	0.00	0.02s
Glass dataset (9 dimensions, 7 clusters)	N	Y(1 component)	7.05	0.60	0.60s
Yeast dataset (8 dimensions, 10 clusters)	Y	Y(7 component)	10.15	0.81	4.77s

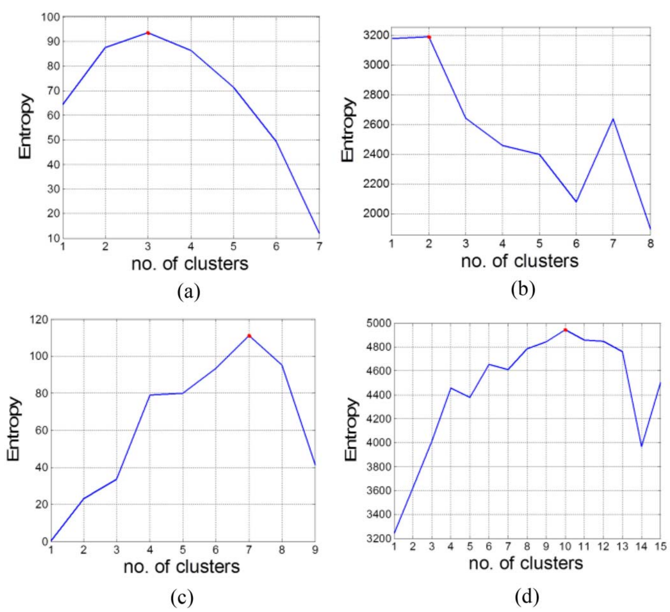

Fig. 14. Entropy criterion of (a) breast dataset, (b) wine dataset, (c) glass dataset, and (d) yeast dataset.

D. Analysis of TSL Words With Entropy-Based *K*-Means Algorithm

To construct the HMM model, we must know how many states should be used in this HMM first. We now apply the entropy-based *K*-means algorithm to analyze the TSL words to figure out the number of states according to the number of clusters.

Fig. 15 gives the entropy-based *K*-means algorithm results of TSL words. Table II tabulates the cluster number for each home-service-related TSL word.

E. HMM Structure and Recognition Result With Cross Validation

Once the number of states for each HMM model is determined, one can construct the classifier, which consists of 11 HMM models. For each HMM model, ABC algorithm is utilized to adjust the structure and the log-likelihood value calculated by Baum–Welch algorithm is applied to determine the best structure for the TSL words. The final structure of HMM is shown in Fig. 16. Fig. 16(a) and (b) shows the "hello" and "good night" structures of HMM, respectively. The "tissue paper" and "coffee" structures are depicted in Fig. 16(c) and (d). Finally, Fig. 16(e) describes the "kitchen" structure of HMM.

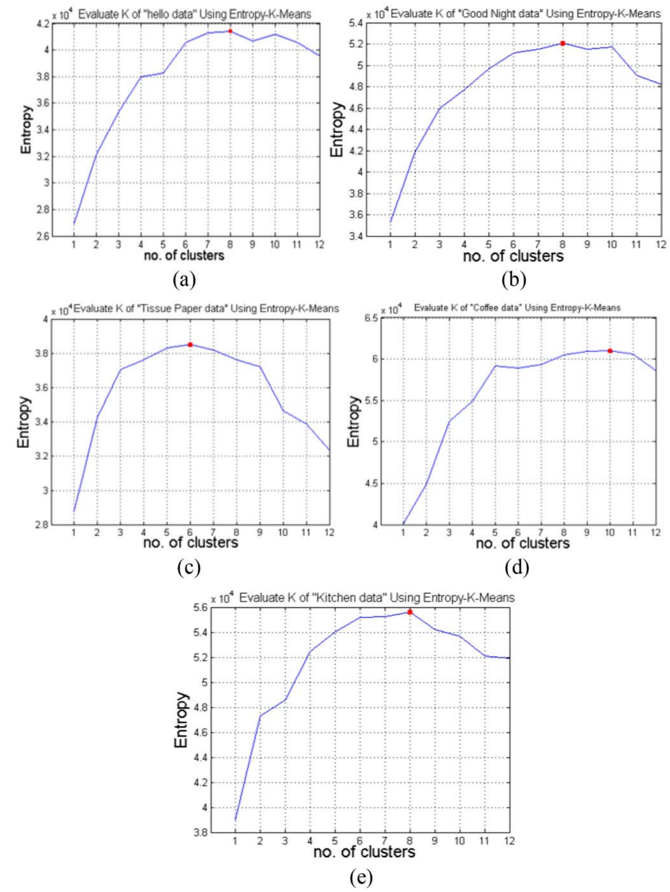

Fig. 15. Entropy-based *K*-means algorithm result of TSL words. (a) Hello. (b) Good night. (c) Tissue paper. (d) Coffee. (e) Kitchen.

TABLE II
CLUSTER NUMBER FOR EACH HOME-SERVICE-RELATED TSL WORD

Entropy-K-means, Distance_scale=0.5, Sample_scale=0.6, Nomralize=Y PCA=2 Component)			
Gesture	No. of Cluster	Gesture	No. of Cluster
Hello	8	Coffee	10
Good Morning	7	Tissue Paper	6
Good Afternoon	6	Living room	7
Good Night	8	Bedroom	6
Milk	6	Kitchen	8
Black Tea	8		

According to the structures learned by the ABC algorithm, one can find that if the TSL word has a repetitive action,

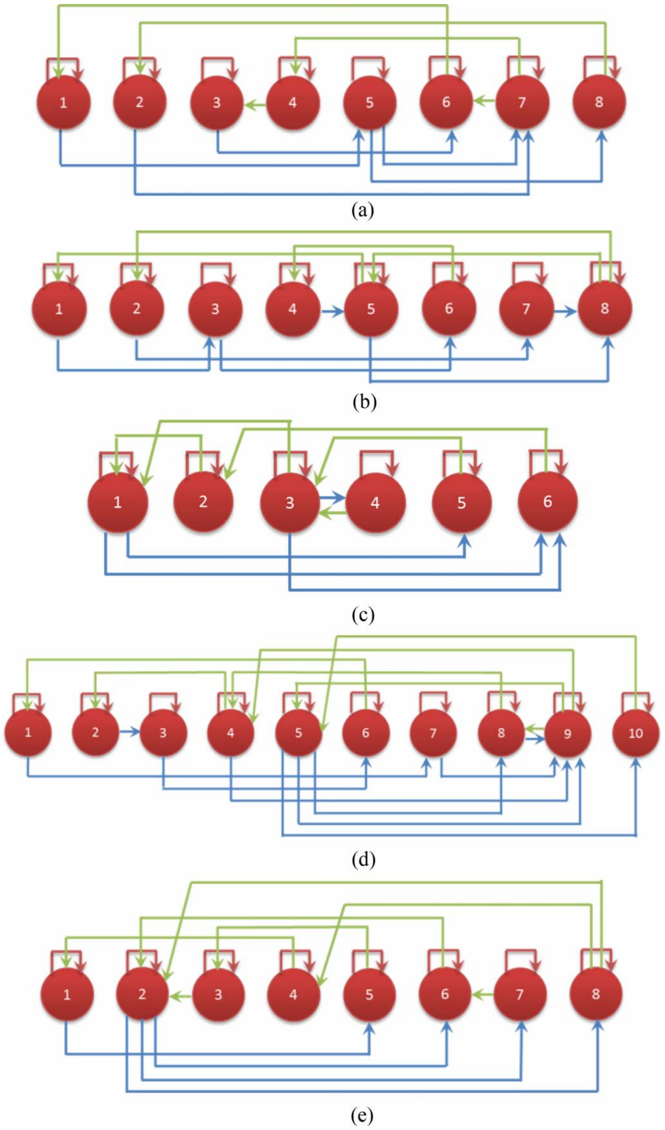

Fig. 16. (a) Hello structure of HMM. (b) Good night structure of HMM. (c) Tissue paper structure of HMM. (d) Coffee structure of HMM. (e) Kitchen structure of HMM.

then the structure will also have a forward and backward structure. For example, tissue paper has a grasping action in front of the nose twice, and the corresponding structure shown in Fig. 16(c) also has a forward and backward behavior between states 3 and 4. Furthermore, coffee has a stirring action with a finger, and the corresponding HMM structure shown in Fig. 16(d) also has a forward and backward behavior between states 8 and 9.

Once the structures of 11 HMM have been determined, we apply the Baum–Welch algorithms to figure out all the parameters of the corresponding HMM. To evaluate the feasibility and effectiveness of the classifier, we adopt the cross-validation method with tenfold. The dataset composed of 11 sign language word data is cut into ten parts randomly, and nine of the parts are used to train the parameters of the classifier; the remaining part is used as the testing data. The process is executed ten times; an average success rate of 91.3% can be achieved.

TABLE III
COMPARATIVE TABLE OF DIFFERENT RECOGNITION ALGORITHMS

Recognition Algorithm	Average Recognition rate	SD
DTW-KNN (K=3) [54]	87.6%	± 3.7%
Left-to-right HMM (8 States) [11][36]	88.2%	± 2.8%
Ergodic HMM (8 states) [53]	82.4%	± 3.8%
Prposed	91.3%	± 1.6%

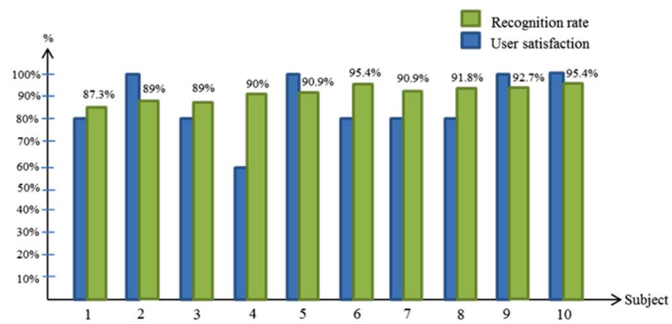

Fig. 17. User satisfaction and recognition rate.

To demonstrate the effectiveness of the proposed model, we have compared our algorithm with left-to-right HMM [11], [36], ergodic HMM [53], and the DTW-KNN method [54]. The experimental results are tabulated in Table III, where the structure of HMM learned by the ABC algorithm is superior to that of the traditional left-to-right HMM and the ergodic HMM. This fact reveals that a suitable structure of HMM contributes to enhancing the recognition rate. References [11] and [36] utilized traditional left-to-right HMM structure to recognize the sign language. They did not apply any methodology to decide the number of states of the HMM model and they just adopted a fixed number of states for all gesture models. Reference [53] employed the ergodic HMM instead of the traditional left-to-right HMM. Although ergodic structure is a new type of structure, it lacks flexibility because all gesture models are fixed on the ergodic structure. In this paper, we employ the data-driven method to learn the structure of each HMM, which can increase the variability of the structure for each gesture model. Moon and Kim [54] utilized DTW with KNN to classify the gestures. Although DTW with KNN also gives a good enough recognition performance, the drawbacks are that the recognition process takes too much time [31], [55] and this method needs a large enough memory to store all the training samples even in the testing phase. However, our method only needs training samples in the training phase to adjust the parameters of the model, so it is not necessary to maintain training samples in the testing phase. Therefore, the memory size and the recognition time can be reduced.

The degree of satisfaction of each subject, i.e., the user satisfaction rate, is also examined and the results are listed in Fig. 17, where most users are satisfied with the recognition system. The recognition rate of subject 1 (who is male) is lower than that of others because he is too nervous to perform the action smoothly. Subject 4 (who is male) thinks that the performance of the system does not meet his expectations; so, the degree of satisfaction is lower than that of others. The recognition rate, however, is not necessarily proportional to the satisfaction. Subject 5 (who is female) feels that the process of the experiments is fun and interesting; so, she is very satisfied with the recognition system.

VI. Conclusion

This paper has proposed an architecture to construct a recognition system for home-service-related TSL words. The proposed recognition model is basically established by HMM and is suitable for modeling time-series data. In general, sign language conveys meaning by utilizing continuous actions that can be considered as a kind of time-sequence data. Although HMM classifier has been utilized for decades, the number of states in each HMM model is decided by the researchers and requires a methodology to determine the number of states. To determine the proper number of states in the HMM, the entropy-based K-means algorithm has been presented to plot the entropy diagram, which provides a visualization method to judge the number of clusters for real datasets. Analyzing the transformed data using the entropy-based K-means algorithm in a different feature space indeed facilitates finding the most suitable feature space to analyze this dataset. The application of the developed entropy-based K-means algorithm to four commonly used datasets has demonstrated its feasibility and validity. Combining the ABC with the Baum–Welch algorithm can successfully establish the data-driven HMM classifier. Finally, this paper has also compared the proposed recognition model with left-to-right HMM, ergodic HMM, and DTW with KNN, and the proposed TSL recognition system can achieve the best correct recognition rate.

References

[1] H.-D. Yang, A.-Y. Park, and S.-W. Lee, "Gesture spotting and recognition for human-robot interaction," *IEEE Trans. Robot.*, vol. 23, no. 2, pp. 256–270, Apr. 2007.

[2] L. Gupta and S. Ma, "Gesture-based interaction and communication: Automated classification of hand gesture contours," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 31, no. 1, pp. 114–120, Feb. 2001.

[3] F. Chersi, "Learning through imitation: A biological approach to robotics," *IEEE Trans. Auton. Mental Develop.*, vol. 4, no. 3, pp. 204–214, Sep. 2012.

[4] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Designing and Implementing a human–robot team for social interactions," *IEEE Trans. Syst., Man, Cybern., Syst*., vol. 43, no. 4, pp. 843–859, Jul. 2013.

[5] Y. Kim and W. C. Yoon, "Generating task-oriented interactions of service robots," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 8, pp. 981–994, Aug. 2014.

[6] A. Aztiria, J. C. Augusto, R. Basagoiti, A. Izaguirre, and D. J. Cook, "Learning frequent behaviors of the users in intelligent environments," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 6, pp. 1265–1278, Nov. 2013.

[7] B. W. Chen, C. Y. Chen, and J. F. Wang, "Smart homecare surveillance system: Behavior identification based on state-transition support vector machines and sound directivity pattern analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 6, pp. 1279–1289, Nov. 2013.

[8] P.-H. Tsai, Y.-J. Lin, Y.-Z. Ou, and E. T.-H. Chu, "A framework for fusion of human sensor and physical sensor data," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 9, pp. 1248–1261, Sep. 2014.

[9] I. Vlasenko, I. Nikolaidis, and E. Stroulia, "The Smart-Condo: Optimizing sensor placement for indoor localization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 3, pp. 436–453, Mar. 2015.

[10] C. H. Wu, Y. H. Chiu, and K. W. Cheng, "Error-tolerant sign retrieval using visual features and maximum a posteriori estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 495–508, Apr. 2004.

[11] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 20, no. 12, pp. 1371–1375, Dec. 1998.

[12] S. Liwicki and M. Everingham, "Automatic recognition of fingerspelled words in British sign language," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 50–57.

[13] I. Infantino, R. Rizzo, and S. Gaglio, "A framework for sign language sentence recognition by commonsense context," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 5, pp. 1034–1039, Sep. 2007.

[14] C. Wang, W. Gao, and S. Shan, "An approach based on phonemes to large vocabulary Chinese sign language recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit*., Washington, DC, USA, 2002, pp. 393–398.

[15] J. S. Kim, W. Jang, and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 2, pp. 354–359, Apr. 1996.

[16] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-temporal feature extraction techniques for isolated gesture recognition in Arabic sign language," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 3, pp. 641–650, Jun. 2007.

[17] Y.-H. Chiu, C.-H. Wu, H.-Y. Su, and C.-J. Cheng, "Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 29, no. 1, pp. 28–39, Jan. 2007.

[18] C.-H. Wu, Y.-H. Chiu, and C.-S. Guo, "Text generation from Taiwanese sign language using a PST-based language model for augmentative communication," *IEEE Trans. Neural Syst. Rehabilit. Eng*., vol. 12, no. 4, pp. 441–454, Dec. 2004.

[19] J. L. Hernandez-Rebollar, N. Kyriakopoulos, and R. W. Lindeman, "A new instrumented approach for translating American sign language into sound and text," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Seoul, Korea, 2004, pp. 547–552.

[20] R. H. Liang and M. Ouhyoung, "A real-time continuous alphabetic sign language to speech conversion VR system," *Comput. Graph. Forum,* vol. 14, no. 3, pp. 67–76, 1995.

[21] W. Gao *et al.*, "HandTalker: A multimodal dialog system using sign language and 3-D virtual human," in *Proc. Int. Conf. Adv. Multimodal Interfaces*, Beijing, China, 2000, pp. 564–571.

[22] N. Habili, C. C. Lim, and A. Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues," *IEEE Trans. Circuits Syst. Video Technol*., vol. 14, no. 8, pp. 1086–1096, Aug. 2004.

[23] Q. Wu, Z. Wang, F. Deng, Z. Chi, and D. D. Feng, "Realistic human action recognition with multimodal feature selection and fusion," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 4, pp. 875–885, Jul. 2013.

[24] K. Assaleh, T. Shanableh, M. Fanaswala, H. Bajaj, and F. Amin, "Vision-based system for continuous Arabic sign language recognition in user dependent mode," in *Proc. 5th Int. Symp. Mechatronics Appl*., Amman, Jordan, 2008, pp. 1–5.

[25] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," in *Proc. Int. Conf. Neural Inf. Process.*, vol. 5. Singapore, 2002, pp. 2204–2206.

[26] G. Fang, W. Gao, and D. Zhao, "Large-vocabulary continuous sign language recognition based on transition-movement models," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 1, pp. 1–9, Jan. 2007.

[27] V. E. Kosmidou, L. J. Hadjileontiadis, and S. M. Panas, "Evaluation of surface EMG features for the recognition of American sign language gestures," in *Proc. IEEE 28th Annu. Int. Conf*. Eng. Med. Biol. Soc., New York, NY, USA, Aug. 2006, pp. 6197–6200.

[28] T. D. Bui and L. T. Nguyen, "Recognizing postures in Vietnamese sign language with MEMS accelerometers," *IEEE Sensors J.*, vol. 7, no. 5, pp. 707–712, May 2007.

[29] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign language recognition using intrinsic mode sample entropy on sEMG and accelerometer data," *IEEE Trans. Biomed. Eng*., vol. 56, no. 12, pp. 2879–2890, Dec. 2009.

[30] X. Zhang *et al.*, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern., A, Syst., Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[31] J. Lichtenauer, E. Hendriks, and M. Reinders, "Sign language recognition by combining statistical DTW and independent classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 2040–2046, Nov. 2008.

[32] M. C. Kao and T. H. S. Li, "Design and implementation of interaction system between humanoid robot and human hand gesture," in *Proc. IEEE/SICE*, Taipei, Taiwan, Aug. 2010, pp. 1616–1621.

[33] C. Alippi, D. Liu, D. Zhao, and L. Bu, "Detecting and reacting to changes in sensing units: The active classifier case," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 3, pp. 353–362, Mar. 2014.

[34] P. Vamplew, "Recognition of sign language gestures using neural networks," in *Proc. Eur. Conf. Disabil. Virtual Real. Assoc. Technol.*, Maidenhead, U.K., 1996, pp. 27–33.

[35] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[36] R. H. Liang and M. Ouhyoung, "A sign language recognition system using hidden Markov model and context sensitive search," in *Proc. ACM Symp. Virtual Real. Softw. Technol.*, Hong Kong, 1996, pp. 59–66.

[37] W. Gao, J. Ma, J. Wu, and C. Wang, "Sign language recognition based on HMM/ANN/DP," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 14, no. 5, pp. 587–602, 2000.

[38] G. Fang, W. Gao, and D. Zhao, "Large vocabulary sign language recognition based on fuzzy decision trees," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 3, pp. 305–314, May 2004.

[39] D. Bouchaffra, "Conformation-based hidden Markov models: Application to human face identification," *IEEE Trans. Neural Netw.*, vol. 21, no. 4, pp. 597–608, Apr. 2010.

[40] L. Antwarg, L. Rokach, and B. Shapira, "Attribute-driven hidden Markov model trees for intention prediction," *J. Syst. Eng. Electron.*, vol. 24, no. 3, pp. 1103–1119, Nov. 2012.

[41] C. H. Wu, J. C. Lin, and W. L. Wei, "Two-level hierarchical alignment for semi-coupled HMM-based audiovisual emotion recognition with temporal course," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1880–1895, Dec. 2013.

[42] H. Li, Y. Yi, X. Li, and Z. Guo, "Human activity recognition based on HMM by improved PSO and event probability," *J. Syst. Eng. Electron.*, vol. 24, no. 3, pp. 545–554, Jun. 2013.

[43] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *Int. J. Comput. Vis.*, vol. 50, no. 2, pp. 203–226, 2002.

[44] D. Malyszko and J. Stepaniuk, "Rough entropy based k-means clustering," in *Proc. 12th Int. Conf. Rough Sets Fuzzy Sets Data Min. Granular Comput.*, Delhi, India, 2009, pp. 406–413.

[45] X. Bai, S. Luo, and Y. Zhao, "Entropy based soft K-means clustering," in *Proc. IEEE Int. Conf. Granular Comput.*, Hangzhou, China, 2008, pp. 107–110.

[46] M. Garcia, H. Hidalgo, and E. Chavez, "Contextual entropy and text categorization," in *Proc. 4th IEEE Comput. Soc. Latin Amer. Web Congr. (LA-WEB)*, Cholula, Mexico, 2006, pp. 147–153.

[47] [Online]. Available: http://cs.joensuu.fi/sipu/datasets/

[48] T. M. Cover and J. A. Thomas, "Entropy, relative entropy and mutual information," in *Elements of Information Theory*. New York, NY, USA: Wiley, 1991, pp. 12–49.

[49] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.

[50] D. Karaboga, "An idea based on bee swarm for numerical optimization," Dept. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. TR06, 2005.

[51] [Online]. Available: http://www.x-io.co.uk/

[52] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabilit. Robot.*, Zurich, Switzerland, Jun. 2011, pp. 1–7.

[53] H.-D. Yang, A.-Y. Park, and S.-W. Lee, "Gesture spotting and recognition for human-robot interaction," *IEEE Trans. Robot.*, vol. 23, no. 2, pp. 256–270, Apr. 2007.

[54] C.-H. Moon and Y.-C. Kim, "Hybrid gesture classifying method using K-NN and DTW for smart remote control," in *Proc. Int. Conf. Inf. Sci. Electron. Elect. Eng. (ISEEE)*, Sapporo, Japan, 2014, pp. 1298–1300.

[55] G. Poli, A. L. M. Levada, J. F. Mari, and J. H. Satio, "Voice command recognition with dynamic time warping (DTW) using graphics processing units (GPU) with compute unified device architecture (CUDA)," in *Proc. 19th Int. Symp. Comput. Archit. High Perform. Comput. (SBACPAD)*, Rio Grande do Sul, Brazil, 2007, pp. 19–25.

Tzuu-Hseng S. Li (S'85–M'90) received the B.S. degree in electrical engineering from the Tatung Institute of Technology, Taipei, Taiwan, in 1981, and the M.S. and Ph.D. degrees in electrical engineering from the National Cheng Kung University (NCKU), Tainan, Taiwan, in 1985 and 1989, respectively.

Since 1985, he has been with the Department of Electrical Engineering, NCKU, where he is currently a Distinguished Professor. From 1996 to 2009, he was a Researcher with the Engineering and Technology Promotion Center, National Science Council, Tainan. From 1999 to 2002, he was the Director of the Electrical Laboratories, NCKU. From 2009 to 2012, he was the Dean of the College of Electrical Engineering and Computer Science, National United University, Miaoli City, Taiwan. He has been the Vice President of the Federation of International Robot-Soccer Association since 2009. His current research interests include artificial and/or biological intelligence and applications, fuzzy system and control, home service robots, humanoid robots, mechatronics, and singular perturbation methodology.

Dr. Li was a recipient of the Outstanding Automatic Control Award in 2006 from the Chinese Automatic Control Society (CACS) in Taiwan. He was a Technical Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS and an Associate Editor of the *Asia Journal of Control*. He is currently an Associate Editor of the *International Journal of Electrical Engineering*, the *International Journal of Fuzzy Systems*, and the IEEE TRANSACTIONS ON CYBERNETICS. He was elected as the President of the CACS from 2008 to 2011 and the Robotics Society of Taiwan from 2012 to 2015. In 2008, he was elevated to CACS Fellow.

Min-Chi Kao received the B.S. degree in electrical science and the M.S. degree in electrical engineering from the National Cheng Kung University (NCKU), Tainan, Taiwan, in 2007 and 2009, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering.

His current research interests include sign language processing, human–machine interaction, fuzzy control, image processing, machine learning, and pattern recognition.

Ping-Huan Kuo was born in Pingtung, Taiwan, in 1986. He received the B.S. and M.S. degrees from the National Cheng Kung University (NCKU), Tainan, Taiwan, in 2008 and 2010, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering.

His current research interests include fuzzy control, intelligent system, humanoid robot, image processing, robotic application, and FIRA/RoboCup game.