2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

A Type-2 Fuzzy Logic Based System for Linguistic Summarization of Video Monitoring in Indoor Intelligent Environments

Bo Yao, Hani Hagras
The Intelligent Environments Group
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
E-mail: byao@essex.ac.uk

Daniyal Alghazzawi, Mohammed J. Alhaddad
King Abdulaziz University
Jeddah, Saudi Arabia

***Abstract*—Video monitoring can provide vital context awareness information from indoor intelligent environments where privacy is not a limitation. However, there is a need to develop linguistic summarization tools which are capable of summarizing in a layman language the information of interest within long video sequences. The key module which can enable the linguistic summarization of video monitoring is human activity/behaviour recognition. However, human behavior recognition is an important yet challenging task due to the behavior uncertainty, activity ambiguity, and uncertain factors such as position, orientation and speed, etc. In order to handle such high levels of uncertainties in activity analysis, we introduce a system based on Interval Type-2 Fuzzy Logic Systems (IT2FLSs) whose parameters are optimized by the Big Bang–Big Crunch (BB-BC) algorithm which allows for robust behaviour recognition using 3D machine vision techniques in intelligent environments. We present several experiments which were performed in real-world intelligent environments to fairly make comparisons with the state-of-the-art algorithms. The experimental results demonstrate that the proposed BB-BC paradigm is effective in tuning the parameters of the membership functions and the rule base of the IT2FLSs to improve the recognition accuracy. It will be shown through real-world experiments that the proposed IT2FLSs outperformed the Type-1 FLSs (T1FLSs) counterpart as well as other traditional non-fuzzy based systems. Based on the recognition results, higher-level applications will presented including video linguistic summarizations event searching and activity retrieval/playback.**

I. INTRODUCTION

The previous years have witnesses an expansion in the installation of video monitoring equipment in public and private spaces. Within indoor intelligent environments where privacy is not a limitation, there is a growing need to develop linguistic summarization tools which are capable of summarizing in a layman language the information of interest within the long video sequences recorded in such spaces. Such summarization can be used to detect automatically serious events that need immediate attention such as attempted burglaries, serious injuries, etc. Linguistic summarization can also provide valuable context information from the video which cannot be extracted by other sensors. For example, an important application in elderly care in intelligent environments is ensuring that the user drinks enough water throughout the day to avoid dehydration. Similarly, a warning message can be sent to social services nearby in case of a fall of an elderly person so that proper actions can be taken instantly. Likewise, lights can be turned off automatically when a user is detected as sleeping on the sofa.

Various works have been attempted for the linguistic summarization of video sequences where type-1 fuzzy logic systems have been applied for linguistic summarization and activities analysis in [1], [2] using voxel person analysis which mainly focuses on fall down detection for eldercare. These type-1 fuzzy based approaches perform well in predefined situations such as a quick change of the silhouette orientation. However, these approaches require time-consuming multi-camera calibration (when one camera is moved slightly, the whole system needs to be re-calibrated). Another linguistic summarization system based on type-1 fuzzy logic was proposed in [3] using wearable devices to summarize and analyze the human activity. However, such wearable devices are intrusive and could be inconvenient for the users.

As a key procedure in linguistic summarization, analysis of human behavior and activity has attracted great interest from researchers. Most previous research on behavior and activity recognition is based on 2D video data [4], [5] or RFID sensors [6]. However, the use of 2D data in real-life circumstances leads to relatively low accuracy due to noise factors and uncertainties associated with real-world environments. The use of RFID tags is intrusive and inconvenient as it requires a deployment of RFID tags on the human or objects. One traditional method is to employ spatio-temporal features to describe points of interest in 2D video data [7]. This method can be later improved by adding more information to model the features [8]. There have been various works for activity analysis employing the Hidden Markov Model (HMM) which was firstly used to analyze two-hand behaviors [9]. Later HMM was used to recognize gesture and pose through a probabilistic framework [10]. However, the accuracy was not satisfactory. Dynamic Time Warping (DTW) is another method to measure similarity and distance between two behaviors [11]. However, DTW only returns exact values and thus is inadequate to model the behavior uncertainty and ambiguity.

In this paper, we present a robust behaviour recognition algorithm for video linguistic summarization using a 3D Kinect camera based on Interval Type-2 Fuzzy Logic

Systems (IT2FLSs) [12], [13]. In order to automatically obtain the optimized parameters of the membership functions and rule base of the IT2FLS, we employed an optimization approach based on the Big Bang–Big Crunch (BB-BC) [22], [23] algorithm. Our experiments have been successfully conducted in real-world intelligent environments and our experiment results show that the proposed IT2FLS outperformed the T1FLS counterpart as well as other traditional non-fuzzy systems. Based on the recognition results, higher-level applications will presented including video linguistic summarizations event searching and activity retrieval/playback.

The rest of the paper is organized as follows. In section II, we provide a brief overview on IT2FLSs. Section III presents an overview on the application of fuzzy logic systems in behavior recognition. Section IV introduces the proposed IT2FLS for the linguistic summarization of video monitoring. Section V presents the automatic optimization approach based on BB-BC for the IT2FLS. Section VI presents the experiments and results. Finally the conclusions and future work are presented in section VII.

II. Brief Overview of Type-2 Fuzzy Logic Systems

The IT2FLS (shown in Fig. 1a) uses the interval type-2 fuzzy sets [12] (shown in Fig. 1b) to represent the inputs and/or outputs of the FLS. In the interval type-2 fuzzy sets all the third dimension values are equal to one [12], [13]. The use of interval type-2 FLS helps to simplify the computation of the type-2 FLS [12].

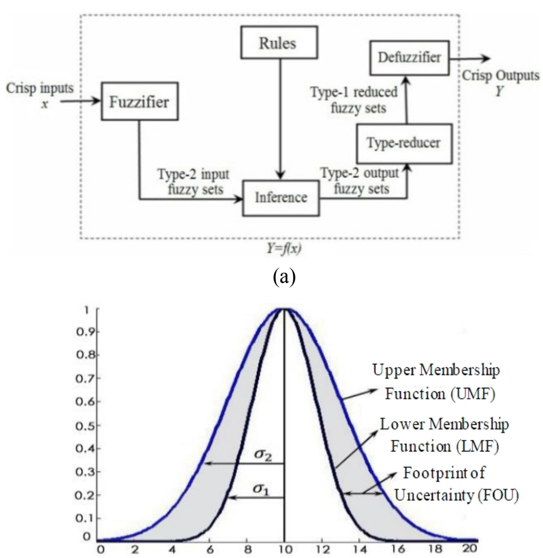


Fig. 1. (a) Structure of the type-2 FLS (b) An interval type-2 fuzzy set.

The interval type-2 FLS works as follows [12], [13]: the crisp inputs from the input sensors are first fuzzified into input type-2 fuzzy sets; singleton fuzzification is usually used in interval type-2 FLS applications due to its simplicity and suitability for embedded processors and real time applications. The input type-2 fuzzy sets then activate the inference engine and the rule base to produce output type-2 fuzzy sets. The type-2 FLS rule base remains the same as for the type-1 FLS but its Membership Functions (MFs) are represented by interval type-2 fuzzy sets instead of type-1 fuzzy sets. The inference engine combines the fired rules and gives a mapping from input type-2 fuzzy sets to output type-2 fuzzy sets. The type-2 fuzzy output sets of the inference engine are then processed by the type-reducer which leads to type-1 fuzzy sets called the type-reduced sets. There are different types of type-reduction methods. In this paper we use the Centre of Sets type-reduction as it has a reasonable computational complexity that lies between the computationally expensive centroid type-reduction and the simple height and modified height type-reductions which have problems when only one rule fires [12], [13]. After the type-reduction process, the type-reduced sets are defuzzified (by taking the average of the type-reduced set) so as to obtain crisp outputs. More information regarding the interval type-2 FLS and its benefits can be found in [12], [13].

III. A Brief Overview of the Application of Fuzzy Logic Systems in Behavior Recognition

Achieving robust behavior and activity recognition in real world environments is highly challenging since even the behaviour features of different subjects which are representative of the same action classes have a wide variance. To make matters worse, the behaviour of a given subject who performs multiple instances of the same action category is not unique. Thus, there are intra- and inter- subject variations in behavioral characteristics which cause high levels of uncertainty in the behaviour recognition.

In [14], a fuzzy rule-based human activity recognition system for e-health was introduced and achieved an accuracy of about 90%. In [15], human activities of a daily living recognition system using hybrid sensors based on fuzzy logic system was proposed and the analysis result was robust. Work in [16] reported an interactive computer graphics environment that encompasses a set of fuzzy logic analysis tools and a fuzzy inference model. In [17], fuzzy logic was employed to recognize students' behavior so as to evaluate their performance in a control course laboratory. In this paper, we employ type-2 fuzzy logic systems in order to handle the faced uncertainties associated with intelligent environments. In [18], a type-2 fuzzy logic based system was employed to learn a user behavior in an intelligent space. However this system used basic sensors and did not employ video or try to generate linguistic summaries of the video sequences.

IV. The Proposed Interval Type-2 Fuzzy Logic System for the Linguistic Summarization of Video Monitoring

Fig. 2 provides an overview of the proposed system. In the learning stage, key pose sequences for each behavior category are recorded. Then behavior features are detected and extracted from Kinect data based on the distance and angle feature information so as to model the motion characteristics. Based on the results of features extraction, the type-1 fuzzy

Membership Functions (MFs) of the inputs to the fuzzy systems are then learned via Fuzzy C-Means Clustering (FCM) [19]. The type-2 fuzzy MFs are then produced by using the obtained type-1 fuzzy sets as the principal membership functions which are then blurred by a certain percentage to create an initial Footprint of Uncertainty (FOU). Finally, the parameters of the IT2FLS are optimized by a method based on BB-BC algorithm.

During the testing stage, input behavior feature vectors are firstly extracted and used as input values for the IT2FLSs-based recognition system. In our fuzzy system, each behaviour model is described by the corresponding rules, and each output degree represents the likelihood between the behaviour in the current frame and the trained behaviour model in the knowledge base. The candidate behaviour in the current frame is then classified and recognized by selecting the candidate model with the highest output degree. Based on the recognition results of the optimized IT2FLS, linguistic summarization is performed using the information of output action category and the duration.

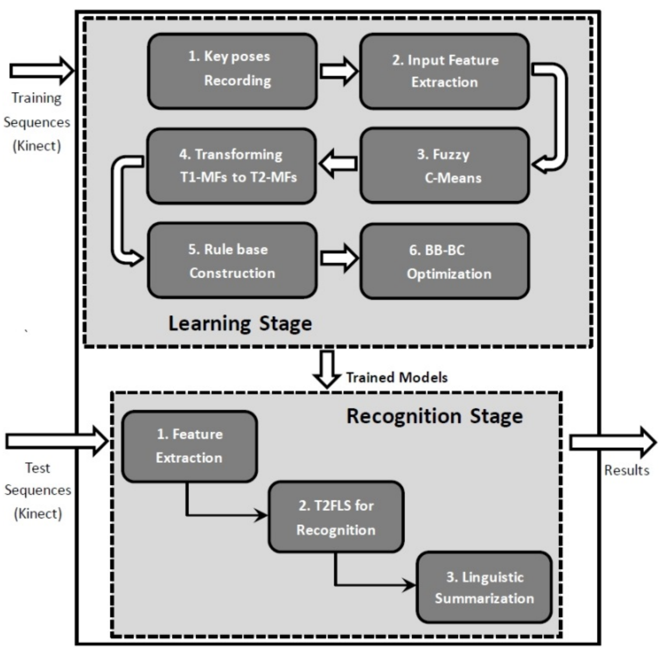


Fig. 2. Overview of our proposed system.

Event retrieval and playback can be conducted after the linguistic summarization, as shown in Fig. 3. As can be seen in the Graphical User Interface (GUI) of our system in Fig. 3, the searching keyword "falldown*" has been inputted and the results of event retrieval are depicted in the list above showing the relevant activities which are detected and stored previously. Both date information and keywords such as behavior categories and duration can be used in our system to search and retrieve the events of interest within a given time frame. In the event searching procedure, input keywords given by the user are compared and matched with the activity information stored in the event base of our system to generate the result list. The retrieved events can be used to playback the video matching the sequences the user wants to see. In the following subsections, we present the various components of the proposed system

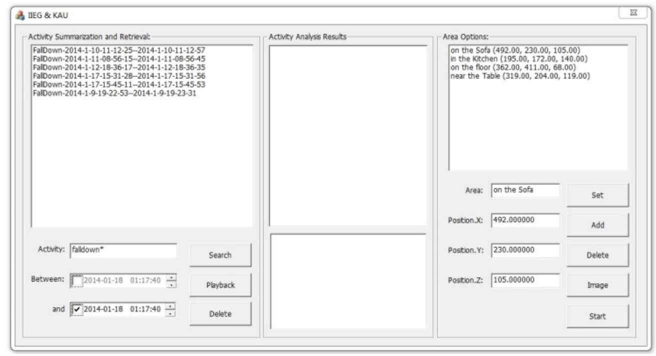


Fig. 3. The GUI of our system

A. *Joint-angle Feature Representation*

For each frame, the skeleton is a sequence of graphs with 15 joints, where each node has its geometric position represented as a 3d point in a global Cartesian coordinate system. An angle feature is defined by three 3D joints P_s, P_o and P_e at a time instant. Firstly, the angle θ is obtained by calculating the angle θ between the vectors $\overrightarrow{P_sP_o}$, and $\overrightarrow{P_oP_e}$ based on the following equation:

$$\theta = cos^{-1}\left(\frac{\overrightarrow{P_sP_o}\times\overrightarrow{P_oP_e}}{\left|\overrightarrow{P_sP_o}\right|\left|\overrightarrow{P_oP_e}\right|}\right) \tag{1}$$

After that, the azimuth angle φ between the vectors $\overrightarrow{P_sO_o}$, and $\overrightarrow{P_oP_e}$ is computed:

B. *Joint-position Feature Representation*

In order to model the local "depth appearance" for the joints, joint position are computed to represent the motion of the skeleton. For distance, between joint i and joint j, the arc-length distance based on equation (2) is calculated:

$$D_{ij} = \left\|P_i - P_j\right\| \tag{2}$$

where $\|\cdot\|$ is the Euclidean norm.

C. *Posture Representation*

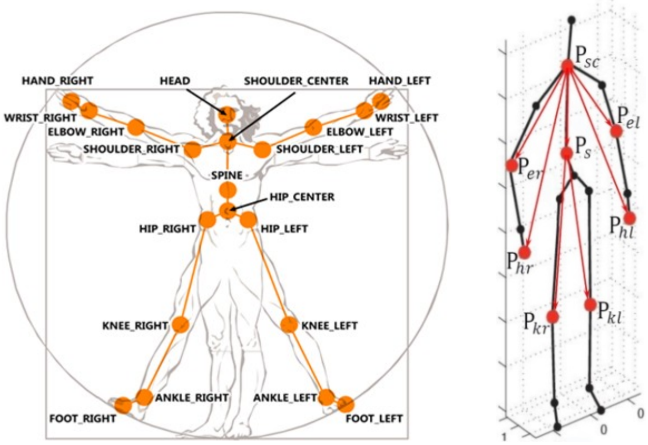
To perform efficient behavior recognition, an appropriate posture representation is essential to model the gesture characteristics so that the similarity and difference can be calculated between current behavior sequence and the recorded key poses. In this work, we use Kinect to extract the 3D skeleton data which comprises 3D joints, shown in Fig. 4a. After that, based on the 3D joints obtained, we compute the posture feature using the joint vectors shown in Fig. 4b.

As most behaviors in daily activity such as drinking, eating, waving hands etc., are related to the upper body, in this work in order to recognize behavior sequence, we focus on the following joints: *shoulder center* (P_{sc}), *elbow left* (P_{el}), *hand left* (P_{hl}), *elbow right* (P_{er}), *hand right* (P_{hr}). Since our algorithm is highly extendable, more joints can be easily added and utilized for more application scenarios. Based on the discussion above, the pose feature is obtained by calculating the joint-angle feature and joint-position feature of the selected joints, as given in the following procedure:

(1) Compute the vectors $\overrightarrow{P_{sc}P_{el}}$, $\overrightarrow{P_{sc}P_{hl}}$ modeling the left arm, and $\overrightarrow{P_{sc}P_{er}}$, $\overrightarrow{P_{sc}P_{er}}$ modeling the right arm.

(2) Angle feature of the left arm θ_{la} can be obtained by calculating the angle between vectors $\overrightarrow{P_{sc}P_{el}}$, $\overrightarrow{P_{sc}P_{hl}}$ based on Equation (1). Then, compute the azimuth angle φ_{la} of the left arm. Similarly, angle features of the right arm θ_{ra} and φ_{ra} can be computed by applying the same process on $\overrightarrow{P_{sc}P_{er}}$, $\overrightarrow{P_{sc}P_{er}}$.

(3) Based on Equation (2), position feature D_{el}, D_{hl}, D_{er}, D_{hr} of the vectors $\overrightarrow{P_{sc}P_{el}}$, $\overrightarrow{P_{sc}P_{hl}}$, $\overrightarrow{P_{sc}P_{er}}$, $\overrightarrow{P_{sc}P_{er}}$ can be obtained.



(a) (b)

Fig. 4. (a) 3D Skeleton and joints. (b) feature vectors.

For each tracked subject at frame *i*, the motion feature vector is obtained:

$$M^i = (\theta^i_{la},\ \varphi^i_{la},\ \theta^i_{ra},\ \varphi^i_{ra},\ D^i_{el},\ D^i_{hl},\ D^i_{er},\ D^i_{hr}) \tag{3}$$

For simplicity, we denote each feature in *M* using the following unique format.

$$M^i = (m^i_0,\ m^i_1,\ m^i_2,\ m^i_3,\ m^i_4,\ m^i_5,\ m^i_6,\ m^i_7) \tag{4}$$

For behavior sequence recognition, key pose recording is performed in the initialization stage by using Equation (3) to model and store the behavior. In our applications, we record the following behaviors: *waving-hands*, *drinking* and *pointing*. Each behavior sequence of key pose is modeled by the motion feature vectors of continuous frames. After key pose recording, we perform training poses for each key pose to model behavior uncertainties such as speed, orientation, and position among others. In this step, DTW [11] is employed to measure the distance between the training pose sequences and the key pose sequences so that the behavior uncertainty and ambiguity can be modeled, as shown in Fig. 5. To achieve this, each feature m^i_j in vector *M* is processed individually by DTW to obtain the distance vector, where *j* = *0*, *1*,..., *7*.

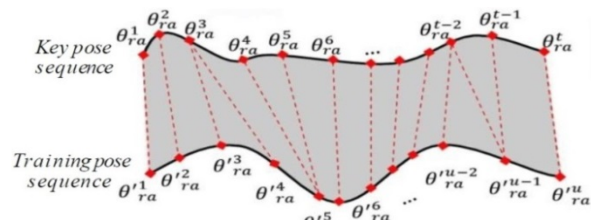


Fig. 5. An example of DTW between key pose sequence and training pose sequence. Corresponding features between the two sequences are indicated by red lines while behavior uncertainty is marked in grey color.

The DTW distance $d_{DTW}(S_{keyPose}, S_{trainingPose})$ between two behavior sequences $S_{keyPose}=\{m^1_j, m^2_j, \ldots, m^t_j\}$ and $S_{trainingPose}=\{m'^1_j, m'^2_j, \ldots, m'^u_j\}$ is defined as:

$$d\ (S_{keyPose}, S_{trainingPose}) = dtw(t,u) \tag{5}$$

$$dtw(v,w) = \min\left\{\begin{matrix} dtw(v-1,w) \\ dtw(v,w-1) \\ dtw(v-1,w-1) \end{matrix}\right\} + d(m^v_j, m'^w_j) \tag{6}$$

where $d(m^v_j, m'^w_j)$ is the Euclidean distance between two poses sequences. Hence, we can obtain the distance vector *V* which is the output result of DTW measuring distance of feature vector *M* between the key pose sequence and training behavior sequence. Thus, vector *V* represents the distance between the key pose sequence and training behavior sequence. The distance vector *V* is shown below, where d_j stands for the distance of the according feature m_j between the key pose and training behavior.

$$V = (d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7) \tag{7}$$

In order to recognize activities including *sitting*, *standing*, *walking*, *running*, and *lying/falling down*, the statue of the spine part of the human subject is calculated which is invariant to orientation and position, as shown below

(1) Compute the vector $\overrightarrow{P_{sc}P_s}$, modeling the spine, and $\overrightarrow{P_{sc}P_{kl}}$, $\overrightarrow{P_{sc}P_{kr}}$ modeling the left knee and right knee.

(2) Compute the angle θ_{kl} between $\overrightarrow{P_{sc}P_s}$ and $\overrightarrow{P_{sc}P_{kl}}$ by using Equation (1). Similarly, the angle θ_{kr} can be obtained by applying Equation (1) on the vectors $\overrightarrow{P_{sc}P_s}$ and $\overrightarrow{P_{sc}P_{kr}}$. Then, the bending angle θ_b of the body can be modeled, which is used mainly for analyzing the *sitting* activity

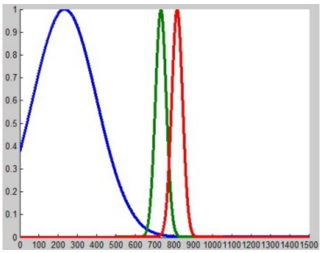
$$\theta_b = (\theta_{kl}+\theta_{kr}) / 2 \tag{8}$$

(3) In order to recognize the *falling down* activity, the 3D coordinates P_{sc} and P_s are firstly projected onto the image plane of the Kinect so as to generate their corresponding 2D coordinates $\tilde{P}_{sc}$ and $\tilde{P}_s$. Then a 2D vector $\overrightarrow{\tilde{P}_{sc}\tilde{P}_s}$ is computed. After that, the lean angle θ_l of the body can be modeled by the orientation of the vector $\overrightarrow{\tilde{P}_{sc}\tilde{P}_s}$, which can be obtained by using the following equation

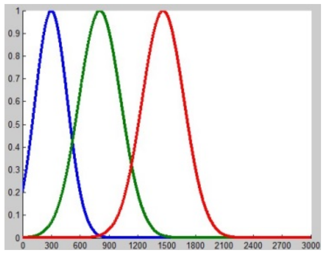
$$\theta_l = tan^{-1}(\frac{y}{x}) \tag{9}$$

where *x* and *y* is the x component and y component of the 2D vector $\overrightarrow{\tilde{P}_{sc}\tilde{P}_s}$.

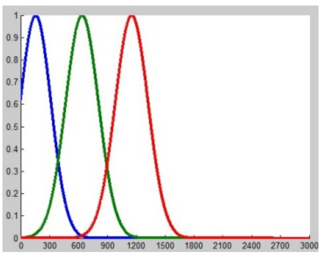
(4) We compute the movement speed of the human by analyzing P_s^{i-1} and P_s^i which are the positions of the joint P_s in two successive frames. The speed D_s can be obtained by applying Equation (1) on P_s^{i-1} and P_s^i. The movement speed D_s is mainly utilized for analyzing the following activities: *falling down, sitting*, *standing*, *walking*, and *running*.



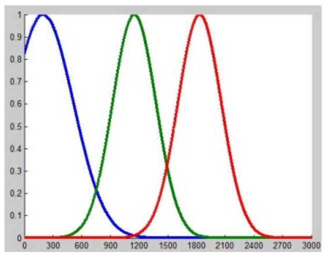
(a) Type-1 MF for d_0



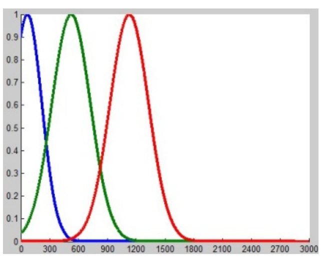
(b) Type-1 MF for d_1



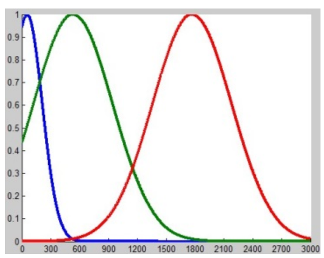
(c) Type-1 MF for d_2



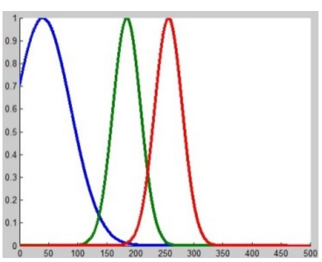
(d) Type-1 MF for d_3



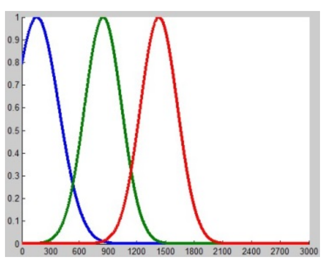
(e) Type-1 MF for d_4



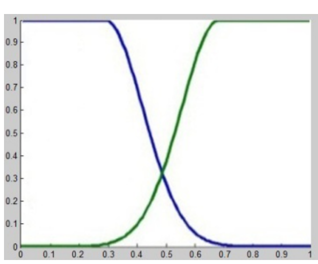
(f) Type-1 MF for d_5



(g) Type-1 MF for d_6



(h) Type-1 MF for d_7



(i) Type-1 MF for Output

Fig. 6. Type-1 membership functions constructed by using FCM.

Based on the above discussion, we can obtain the feature vector for activity analysis

$$A = (\theta_b, \theta_l, D_s) \tag{10}$$

D. The Type-1 Fuzzy Logic System for Behavior and Activity Recognition

It should be noted that we present here the T1FLS which will be used for comparison against our IT2FLS. Both feature vectors V and A will be analyzed by of FLS in the same paradigm but separately where the feature vector V will be employed for behavior recognition while vector A will be employed for activity analysis.

In order to recognize behavior sequence based on the feature vector V, in our fuzzy system, the antecedents are d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7. Each of these antecedents is represented by three fuzzy sets which are *LOW*, *MEDIUM*, and *HIGH*. The output of the fuzzy system is the behaviour possibility which is represented by two fuzzy sets which are *LOW* and *HIGH*. The type-1 MFs shown in Fig. 6 have been obtained via FCM, and are then transformed to type-2 MFs

Suppose we measure $\{d_0, d_1, ..., d_7\}$ on current behavior of the 3D skeleton expressing the possibilities of the candidate behaviour classes: *drinking*, *waving-hands*, and *pointing*. The mapping between measurement and behaviour classes is accomplished by fuzzy rules. In our system, the rule base is constructed by using the Wang-Mendel approach [13], [20], [21].

Each behaviour class uses the same output membership function that is shown in Fig. 6i. In the T1FLS, product *t*-norm is used to represent the AND logical connective and the implication operation. The behaviour recognition is conducted via selecting the best candidate behaviour class with the highest output as the recognized behaviour type. However, if two different candidate behaviour classes are assigned with the same output degree, this means that these two candidate behaviour classes have significantly high behavioral similarity and cannot be distinguished in the current frame.

E. Transforming Type-1 Membership Functions to Interval Type-2 Membership Functions

In this subsection, we present the manual design process of the IT2FLS which will be further optimized by the proposed BB-BC algorithm presented in the next section. In the experiment section, we compare the results obtained by this manually designed IT2FLS against the IT2FLS optimized by BB-BC.

The interval type-2 fuzzy set Footprint of Uncertainty (FOU) is bounded by two MFs which are the Upper Membership Function (UMF) and the Lower Membership Function (LMF), respectively. As shown in Fig. 7, for example, the fuzzy set HIGH is represented by the type-2 Gaussian membership function whose mean and standard deviation are obtained from numerous measurements of the azimuth angle of left arm φ_{la}.

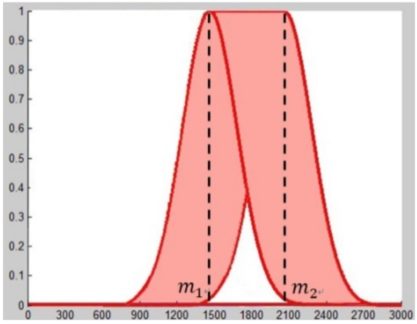


Fig. 7. Gaussian MF of numerous instances of feature left arm azimuth angle θ_{la} for the linguistic variable HIGH

In order to construct the type-2 MFs modeling the FOU, we transform the type-1 fuzzy sets to the interval type-2 fuzzy sets with uncertain mean. We consider the case of a Gaussian primary membership function having a fixed standard deviation σ and an uncertain mean m that in the range $[m_{k1}^l, m_{k2}^l]$ [13], i.e.,

$$u_k^l(x_k) = \exp\,[-\tfrac{1}{2}\left(\tfrac{x_k-m_k^l}{\sigma_k^l}\right)]\ ,\ m_k^l \in [m_{k1}^l, m_{k2}^l] \tag{11}$$

where $k = 1,\ldots, p$; p is the number of antecedents; $l = 1,\ldots, M$; M is the number of rules. The upper membership function of the type-2 fuzzy set can be written as follows:

$$\bar{u}_k^l(x_k) = \begin{cases} N\left(m_{k1}^l, \sigma_k^l, x_k\right), & x_k < m_k^l \\ 1, & m_{k1}^l \le x_k \le m_{k2}^l \\ N\left(m_{k2}^l, \sigma_k^l, x_k\right), & x_k > m_{k2}^l \end{cases} \tag{12}$$

The lower membership function can be written as follows:

$$\underline{u}_k^l(x_k) = \begin{cases} N\left(m_{k2}^l, \sigma_k^l, x_k\right), & x_k \le \frac{m_{k1}^l + m_{k2}^l}{2} \\ N\left(m_{k1}^l, \sigma_k^l, x_k\right), & x_k > \frac{m_{k1}^l + m_{k2}^l}{2} \end{cases} \tag{13}$$

where

$$N\left(m_{k1}^l, \sigma_k^l, x_k\right) = \exp\left(-\frac{1}{2}\left(\frac{x_k-m_{k1}^l}{\sigma_k^l}\right)\right) \tag{14}$$

In order to construct the type-2 MFs in our IT2FLS, we use the mean of the given type-1 fuzzy set (extracted by FCM clustering in the previous subsection) to represent the m_{k1}^l, m_{k2}^l is obtained by blurring m_{k1}^l with a certain α% (α = *10*, *20*, *30*, *40*…) such that

$$m_{k2}^l = (1+\alpha\%)\, m_{k1}^l \tag{15}$$

where σ_k^l is the same as the given type-1 fuzzy set. In order to allow for a fair comparison between the type-2 fuzzy logic system and type-1 fuzzy logic system, we used the same input features for the IT2 FLS and the T1FLS.

V. The Proposed Optimization Method for the IT2FLS

In order to optimize the proposed IT2FLS, the MFs and the rule base have to be determined. In this proposed system, the BB-BC is utilized to calculate the optimized parameters for the MFs and the rules of our IT2FLS.

A. *Big Bang-Big Crunch (BB-BC) Optimization*

The BB-BC optimization is a heuristic population based on evolutionary approach which was presented by Erol and Eksin [22]. It is derived from one of the theories of the evolution of universe in physics and astronomy, namely the BB-BC theory. The key advantages of BB-BC are its low computational cost, ease of implementation, and fast convergence. The BB-BC theory is formed by two phases: a Big Bang phrase where candidate solutions are randomly distributed over the search space in a uniform manner [23] and a Big Crunch phrase where candidate solutions are drawn into a single representative point via a center of mass or minimal cost approach [22]. All subsequent Big Bang phases are randomly distributed around the center of mass or the best fit individual in a similar fashion. The procedures followed in the BB-BC are as follows [23]:

Step 1 (Big Bang Phase): An initial generation of N candidates is randomly generated in the search space, similar to the other evolutionary search algorithms.

Step 2: The cost function values of all the candidate solutions are computed.

Step 3 (Big Crunch Phase): Big Crunch phase comes as a convergence operator. Either the best fit individual or the center of mass is chosen as the center point. The center of mass is calculated as:

$$x_c = \frac{\sum_{i=1}^{N} \frac{x_i}{f^i}}{\sum_{i=1}^{N} \frac{1}{f^i}} \tag{16}$$

where x_c is the position of the center of mass, x_i is the position of the candidate, f^i is the cost function value of the i^{th} candidate, and N is the population size.

Step 4: New candidates are calculated around the new point calculated in Step *3* by adding or subtracting a random number whose value decreases as the iterations elapse, which can be formalized as:

$$x^{new} = \beta x_c + (1-\beta) x_{best} + \frac{\gamma\rho(x_{max}-x_{min})}{k} \tag{17}$$

where β is the parameter controlling the influence of the global best solution x_{best} on the location of new candidate solutions, r is random number, ρ is a parameter limiting search space, x_{min} and x_{max} are the lower and upper limits.

Step 5: Return to Step *2* until stopping criteria have been met.

B. Optimizing the Type-2 membership functions with BB-BC

In order to apply BB-BC, the feature parameters of the type-2 MFs have to be encoded into a form of a population. As depicted in Equation (15), in order to construct the type-2 MFs, the parameter α has to be determined to obtain m_{k2}^{l} while m_{k1}^{l} is provided by FCM. To be more accurate, the uncertainty factors α_k^j for each fuzzy set of the MFs are computed, where $k = 1,\ldots, p$, p is the number of antecedents; $j = 1,\ldots, q$, q is the number of input features. For illustration purposes, as in the MFs of proposed system, three type-2 fuzzy sets including *LOW, MEDIUM and HIGH* are utilized for modeling each of the 8 features, therefore, the total number of the parameters for the input type-2 MFs is 3×8=24. In similar manner, parameters the output MFs are also encoded; these are α_L^{Out} for the linguistic variable LOW and α_H^{Out} for the linguistic variable HIGH of the output MF. Therefore, the structure of the population is built as displayed in Fig. 8.

α_1^0	α_2^0	α_3^0	...	α_1^7	α_2^7	α_3^7	α_L^{Out}	α_H^{Out}

Fig. 8. The population representation for the parameters of type-2 MFs.

The optimization problem is a minimization task, and with the parameters of the MFs encoded as showed in Fig. 8 and the constructed rule base, the recognition error in our solutions space can be minimized by using the following function as the cost function for BB-BC.

$$f^i = (1 - Accuracy^i) \tag{18}$$

where f^i is the cost function value of the i^{th} candidate and $Accuracy^i$ is the scaled recognition accuracy of the i^{th} candidate.

C. Optimizing the rule base of the IT2FLS with BB-BC

In similar manner as optimizing the MFs using BB-BC, the parameters of the rule base are encoded into a form of a population. The IT2FLS rule base can be represented as shown in Fig. 9.

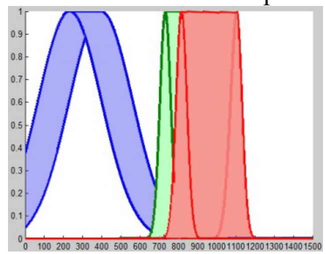
d_0^1	d_2^1	...	d_7^1	d_{out}^1	...	d_0^R	d_2^R	...	d_7^R	d_{out}^R

Fig. 9. The population representation for the parameters of type-2 MFs.

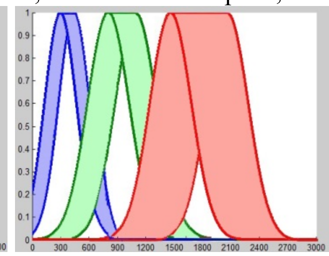
As showed in Fig. 9, d_k^r are the antecedents and d_{out}^r is the consequent of each rule respectively, where $k = 1,\ldots, p$, p is the number of antecedents; $r = 1,\ldots, R$, R is the number of the rules to be tuned. In this study, the rule base constructed by the Wang-Mendel approach [13], [20], [21] is used as the initial generation of candidates. After that, the rule base can be tuned by BB-BC using the cost function depicted in equation 18.

VI. Experiments and Results

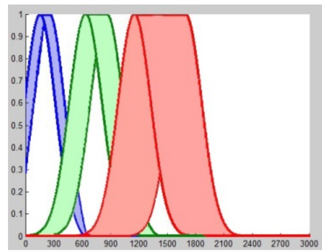
We tested both of our T1FLS and IT2FLS in the iSpace located in the University of Essex which is a real-world intelligent environment. The experiments were conducted with different subjects and different scenes in various circumstances including different illumination conditions, daytime and nighttime, fixed-camera and moving camera, etc. The experiment results demonstrate that our algorithm is robust and effective in handling the high levels of uncertainties associated with real-world environments including behavior uncertainty, activity ambiguity, and uncertain factors such as position, orientation and speed, etc.



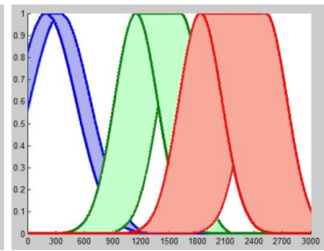
(a) Type-2 MF for d_0



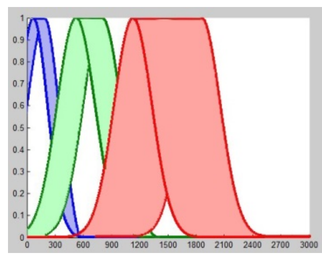
(b) Type-2 MF for d_1



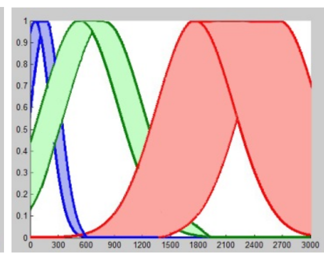
(c) Type-2 MF for d_2



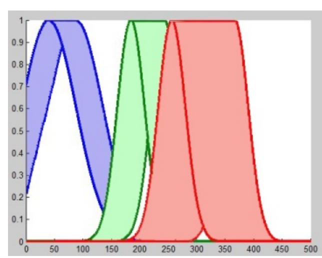
(d) Type-2 MF for d_3



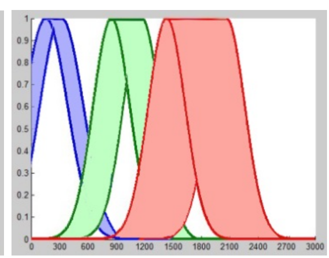
(e) Type-2 MF for d_4



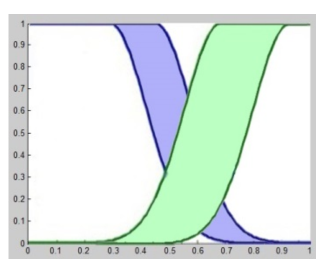
(f) Type-2 MF for d_5



(g) Type-2 MF for d_6



(h) Type-2 MF for d_7



(i) Type-2 MF for Output

Fig. 10. Type-2 membership functions optimized by using BB-BC

The type-2 membership functions used in our system which are obtained and optimized by BB-BC are shown in

Fig. 10. Our experiment result demonstrates that the BB-BC optimization improves the performance of our type-2 fuzzy logic system. In the BB-BC optimization procedure of the type2 membership functions, we set x_{min} and x_{max} to 40% and 60%, which are regarding to the FOU blurring factor α in type-2 MFs construction. In order to achieve robust recognition performance, in our experiment, the population size N of BB-BC is set to 100000. And owing to the high-performance of BB-BC, each iteration of the optimization procedure can be done in a few minutes.

Based on the optimized type-2 MFs and rule base by utilizing BB-BC, our IT2FLSs-based system outperforms the counterpart T1FLSs-based recognition system, as shown in Table I, the type-2 system without BB-BC achieves 2.63% higher average per-frame accuracy than the T1FLS. In addition, the BB-BC optimized IT2FLS achieves 3.72% higher average per-frame accuracy than the BB-BC optimized T1FLS.

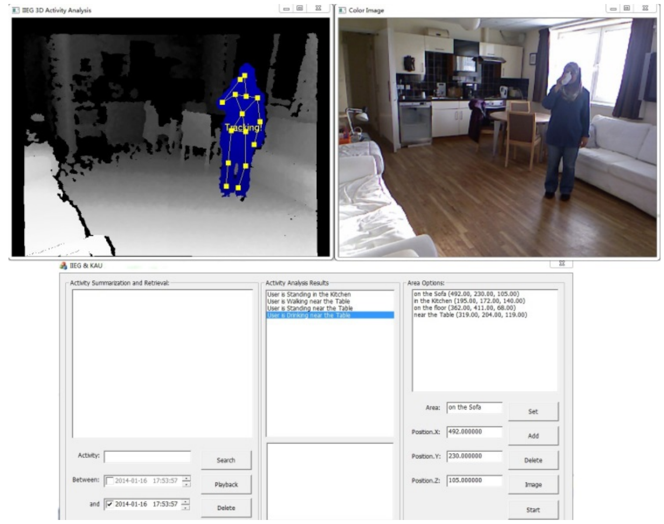
Our type-2 fuzzy logic system also outperforms the traditional non-fuzzy based recognition method which used dynamic time warping (DTW) [11]. In order to conduct a fair comparison with the traditional DTW-based method, our IT2FLSs-based system utilizes the same input features with the DTW-based method. As shown in Table I, our IT2FLSs-based method with BB-BC optimization achieves 7.97% higher recognition average accuracy than the DTW-based algorithm.

Based on the recognition results of our optimized IT2FLS, higher-level applications including video linguistic summarizations, event searching, activity retrieval, event playback, and human-machine interactions have been developed and successfully deployed in our real-world test bed iSpace [18], [21] which is the ambient intelligent environment in the University of Essex

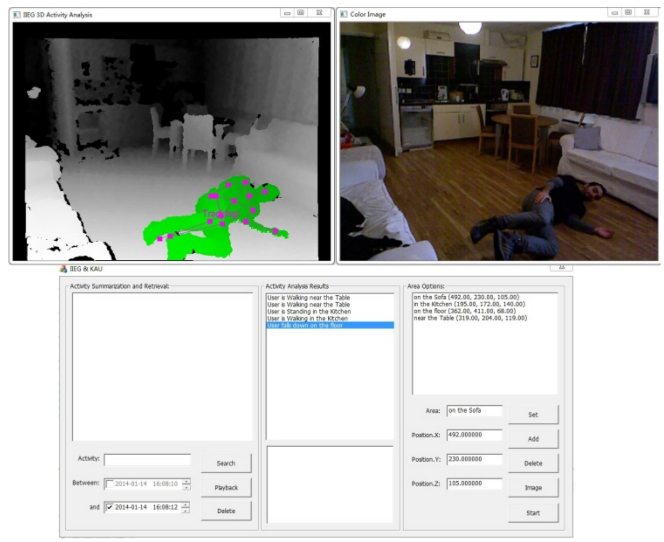
TABLE I
COMPARISON OF OVERALL AVERAGE ACCURACY FOR BEHAVIOR RECOGNITION WITH PREVIOUS TRADITIONAL NON-FUZZY METHOD

Methods	Recognition Accuracy
IT2FLSs with BB-BC	97.48%
IT2FLSs without BB-BC	96.35%
T1FLSs-based method	93.72%
Dynamic Time Warping [11]	89.51%

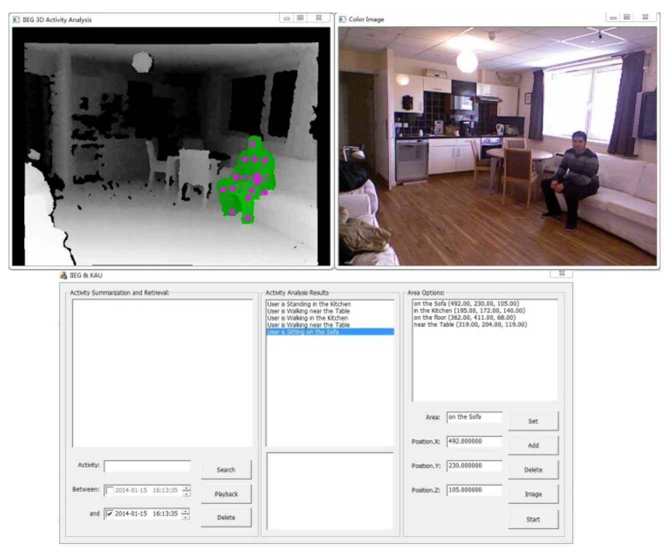
As can be seen in Fig. 11, both of the original images and depth images captured by Kinect are shown above the main GUI of our system. And in the depth images, human is detected and highlighted with a specified color together with the 3D skeleton. And the current analysis result is depicted and marked in color in the middle list of the main GUI. As can be seen in Fig. 11d, the user can easily summarize the event of interest at the given time frame and play them back.



(a) Recognition results for Behavior "Drinking"



(b) Recognition results for Behavior "Fall down"



(c) Recognition results for Behavior "Sitting"

(d) Linguistic summarization and playback based on given time frame

Fig. 11. Real-time analysis results by our IT2FLS-based system

VII. Conclusions and Future Work

In this paper, we presented a robust behaviour recognition algorithm for video linguistic summarization using a 3D Kinect camera based on Interval Type-2 Fuzzy Logic Systems (IT2FLSs). In order to automatically obtain the optimized parameters of the membership functions and rule base of the IT2FLS, we employed an optimization approach based on the Big Bang–Big Crunch (BB-BC) [22], [23] algorithm. Our experiments have been successfully conducted in real-world intelligent environments and our experiment results show that the proposed IT2FLS outperformed the T1FLS counterpart as well as other traditional non-fuzzy systems. Based on the recognition results, higher-level applications were presented including video linguistic summarizations event searching and activity retrieval/playback.

In our future ongoing research, we intend to develop higher-level specific applications such as elder/children care, event surveillance and summarization and human-machine interaction based on the proposed IT2FLS-based system of human behaviour recognition and apply the applications in real-world intelligent indoor and outdoor environments so as to construct the advanced scale-up intelligent environments.

References

[1] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, “Linguistic summarization of video for fall detection using voxel person and fuzzy logic” Comput. Vision Image Understanding, vol. 113, pp. 80–89, 2008.

[2] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, “Modeling human activity from voxel person using fuzzy logic” IEEE Transactions on Fuzzy Systems, vol. 17, no. 1, pp. 39-49, 2009.

[3] G. Trivino, and A. van der Heide, “Linguistic summarization of the human activity using skin conductivity and accelerometers” 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU), pp. 1583-1589, 2008.

[4] J. K. Aggarwal, and M. S. Ryoo, “Human activity analysis: A review” ACM Computing Surveys (CSUR), vol. 43, no. 3, pp. 16, 2011.

[5] M. S. Ryoo, and J. K. Aggarwal, “Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities” 12th IEEE International Conference on Computer Vision, pp. 1593-1600, 2009.

[6] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, “A scalable approach to activity recognition based on object use” 11th IEEE International Conference on Computer Vision, ICCV 2007, pp. 1-8, 2007.

[7] I. Laptev, “On space-time interest points” International Journal of Computer Vision, vol. 64, no. 2-3, pp.107-123, 2005.

[8] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, “A biologically inspired system for action recognition” 11th IEEE International Conference on Computer Vision, ICCV 2007, pp. 1-8, 2007.

[9] M. Brand, N. Oliver, and A. Pentland, “Coupled hidden Markov models for complex action recognition” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 994-999, 1997.

[10] A. Elgammal, V. Shet, Y. Yacoob, and L. S. Davis, “Learning dynamics for exemplar-based gesture recognition” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-571, 2003.

[11] F. Zhou, F. Torre, and J. K. Hodgins, “Aligned cluster analysis for temporal segmentation of human motion” 8th IEEE International Conference on Automatic Face & Gesture Recognition, FG'08., pp. 1-7. IEEE, 2008.

[12] H. Hagras, “A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots” IEEE Transactions on Fuzzy Systems, vol. 12, no. 4, pp.524-539, 2004.

[13] J. Mendel, “Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions” Upper Saddle River, NJ: Prentice-Hall, 2001.

[14] J. Y. Chang, J. J. Shyu, and C. W. Cho, “Fuzzy rule inference based human activity recognition” Control Applications,(CCA) & Intelligent Control,(ISIC), IEEE, pp. 211-215, 2009.

[15] H. Medjahed, D. Istrate, J. Boudy, and B. Dorizzi, “Human activities of daily living recognition using fuzzy logic for elderly home monitoring” IEEE International Conference on Fuzzy Systems, pp. 2001-2006, FUZZ-IEEE, 2009.

[16] I. A. Ioannidou, S. Paraskevopoulos, and P. Tzionas, “An interactive computer graphics interface for the introduction of fuzzy inference in environmental education” Interacting with Computers, vol. 18, no. 4, pp. 683-708, 2006.

[17] G. Gokmen, T. Ç. Akinci, M. Tektaş, N. Onat, G. Kocyigit, and N. Tektaş, “Evaluation of student performance in laboratory applications using fuzzy logic” Procedia-Social and Behavioral Sciences, vol. 2, no. 2, pp. 902-909, 2010.

[18] H. Hagras, F. Doctor, A. Lopez, V. Callaghan, “An Incremental Adaptive Life Long Learning Approach for Type-2 Fuzzy Embedded Agents in Ambient Intelligent Environments” IEEE Transactions on Fuzzy Systems, vol. 15, no. 1, pp. 41-55, February 2007.

[19] N. R. Pal and J. C. Bezdek, “On cluster validity for the fuzzy c-means model” IEEE Transaction on Fuzzy Systems, vol. 3, pp. 370-379, 1995.

[20] L. X. Wang, “The MW method completed: A flexible system approach to data mining” IEEE Transactions on Fuzzy Systems, vol. 11, no. 6, pp. 768–782, December 2003.

[21] H. Hagras, F. Doctor, A. Lopez and V. Callaghan, “An incremental adaptive life long learning approach for type-2 fuzzy embedded agents in ambient intelligent environments” IEEE Transactions on Fuzzy Systems, vol. 15, no. 1, pp. 41-55, 2007.

[22] O. Erol and I. Eksin, “A new optimization method: big bang–big crunch” Advances in Engineering Software, vol.37, no. 2, pp. 106-111, 2006.

[23] T. Kumbasar, I. Eksin, M. Guzelkaya and E. Yesil, “Adaptive fuzzy model based inverse controller design using BB-BC optimization algorithm” Expert Systems with Applications, vol. 38, no. 10, pp.12356-12364, 2011.