A Big Bang Big Crunch Type-2 Fuzzy Logic System for Machine Vision-Based Event Detection and Summarization in Real-world Ambient Assisted Living

Bo Yao

A thesis submitted for the degree of Doctor of Philosophy in Computer Science

School of Computer Science and Electronic Engineering
University of Essex
April 2015

Abstract

The recent years have witnessed the prevalence and abundance of vision sensors in various applications such as security surveillance, healthcare and Ambient Assisted Living (AAL) among others. This is so as to realize intelligent environments which are capable of detecting users' actions and gestures so that the needed services can be provided automatically and instantly to maximize user comfort and safety as well as to minimize energy. However, it is very challenging to automatically detect important events and human behaviour from vision sensors and summarize them in real time. This is due to the massive data sizes related to video analysis applications and the high level of uncertainties associated with the real world unstructured environments occupied by various users. Machine vision based systems can help detect and summarize important information which cannot be detected by any other sensor; for example, how much water a candidate drank and whether or not they had something to eat. However, conventional non-fuzzy based methods are not robust enough to recognize the various complex types of behaviour in AAL applications.

Fuzzy logic system (FLS) is an established field of research to robustly handle uncertainties in complicated real-world problems. In this thesis, we will present a general recognition and classification framework based on fuzzy logic systems which allows for behaviour recognition and event summarisation using 2D/3D video sensors in AAL applications. I started by investigating the use of 2D CCTV camera based system where I proposed and developed novel IT2FLS-based methods for silhouette extraction and 2D behaviour recognition which outperform the traditional on the publicly available Weizmann human action dataset. I will also present a novel system

based on 3D RGB-D vision sensors and Interval Type-2 Fuzzy Logic based Systems (IT2FLSs)) generated by the Big Bang Big Crunch (BB-BC) algorithm for the real time automatic detection and summarization of important events and human behaviour. I will present several real world experiments which were conducted for AAL related behaviour with various users. It will be shown that the proposed BB-BC IT2FLSs outperforms its Type-1 FLSs (T1FLSs) counterpart as well as other conventional non-fuzzy methods, and that performance improvement rises when the number of subjects increases. It will be shown that by utilizing the recognized output activity together with relevant event descriptions (such as video data, timestamp, location and user identification) detailed events are efficiently summarized and stored in our back-end SQL event database, which provides services including event searching, activity retrieval and high-definition video playback to the front-end user interfaces.

Acknowledgments

I take this opportunity to express my most sincere gratitude and appreciation to my supervisor Professor Hani Hagras, who has given me continuous guidance and encouragement throughout these years. He has offered me many critical opportunities that have changed my life and deeply benefited my career. He has also inspired me by keeping my motivation high and cultivating my enthusiasm towards science, innovation and research.

Furthermore, I would like to express my appreciation to Dr Mohamed Al Haddad, Dr Daniyal Alghazzawi and Dr Steel Sam for their encouragement and insightful comments. I would also like to thank my wife Xiao Yan and my parents Tianquan Yao and Zuhui Zhou who have constantly and unconditionally supported me throughout the entire tough process of my Ph.D. research.

Last but not least, I would also like to specially thank King Abdulaziz University for sponsoring my Ph.D. research and providing me valuable teaching and consulting opportunities from which I benefited greatly.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAL	Ambient Assisted Living
BB-BC	Big Bang Big Crunch
FLS	Fuzzy Logic System
T1FLS	Type-1 Fuzzy Logic System
T2FLS	Type-2 Fuzzy Logic System
IT2FLS	Interval Type-2 Fuzzy Logic System
WHO	World Health Organization
CCTV	Closed-Circuit Television
RFID	Radio-frequency identification
CRF	Conditional Random Fields
HMM	Hidden Markov Model
DTW	Dynamic Time Warping
2D	Two-dimensional
3D	Three-dimensional
GMM	Gaussian Mixture Model
GUI	Graphical User Interface
UMF	Upper Membership Function

LMF	Lower Membership Function
FOU	Footprint of Uncertainty
FS	Fuzzy Set
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
GA	Genetic Algorithm
MF	Membership Function
T1MF	Type-1 Fuzzy Membership Function
T2MF	Type-2 Fuzzy Membership Function
RGB	Red Green Blue
RGB-D	Red Green Blue-Depth
***k*-NN**	*k*-Nearest Neighbour
SVM	Support Vector Machine
FCM	Fuzzy C-Means
MEIs	Motion Energy Images
MHIs	Motion History Images
STIPs	Spatio-Temporal Interest Points
HOG	Histogram of Oriented Gradients
HOF	Histogram of Optic Flows

GNN	Global Nearest Neighbour
SAD	Sum of Absolute Difference)
hCRF	Hidden Conditional Random Field
HD	High-Definition
SDK	Software Development Kit
OpenNI	Open Natural Interaction
PCL	Point Cloud Library
CUDA	Compute Unified Device Architecture
SQL	Structured Query Language
COS	Center of Sets
PDF	Probability Density Function
BOPs	Bag of 3D Points
DMF	Deformable Model Fitting
iClassroom	Intelligent Classroom
iSpace	Intelligent Space

Publications Arising from This Work

1. **Bo Yao**, Hani Hagras, Mohammed J. Alhaddad, and Daniyal Alghazzawi, “A Big Bang Big Crunch Type-2 Fuzzy Logic System for Machine Vision-Based Event Detection and Summarization in Real-world Ambient Assisted Living,” submitted to IEEE Transaction on Fuzzy Systems, 2015.

2. Khalid Almohammadi, **Bo Yao**, and Hani Hagras, “An Interval Type-2 Fuzzy Logic Based System for Improved Instruction within Intelligent E-Learning Platforms,” accepted by 2015 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2015, Istanbul, Turkey, August 2015.

3. **Bo Yao**, Hani Hagras, Mohammed J. Alhaddad, and Daniyal Alghazzawi, “A Type-2 Fuzzy Logic System for Linguistic Summarization of Video Sequence in Indoor Intelligent Environments,” IEEE World Congress on Computational Intelligence, pp. 825-833, IEEE WCCI 2014, Beijing, China, July 2014 (Won the **Best Student Paper Award in FUZZ-IEEE 2014**).

4. Khalid Almohammadi, **Bo Yao**, and Hani Hagras, “An Interval Type-2 Fuzzy Logic Based System with User Engagement Feedback for Customized Knowledge Delivery within Intelligent E-Learning Platforms,” IEEE World Congress on Computational Intelligence, pp. 808-817, IEEE WCCI 2014, Beijing, China, July 2014.

5. **Bo Yao**, Hani Hagras, Mohammed J. Alhaddad, and Daniyal Alghazzawi, "A Fuzzy Logic Based System for the Automation of Human Behavior Recognition using Machine Vision in Intelligent Environments," Journal of Soft Computing, vol. 19, pp. 499-506, 2014.

6. **Bo Yao**, Hani Hagras, Danniyal Al Ghazzawi, and Mohammed J. Alhaddad, "Type-2 Fuzzy Logic Machine Vision Based Approach for Human Behaviour Recognition in Intelligent Environments," IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'13, Hyderabad, India, pp. 1-8, 2013.

7. **Bo Yao**, Hani Hagras, Daniyal Al Ghazzawi, and Mohammed J. Alhaddad, "A Big Bang-Big Crunch Optimization for a Type-2 Fuzzy Logic based Human Behaviour Recognition System in Intelligent Environments," IEEE International Conference on Systems, Man, and Cybernetics, SMC'13, pp. 2880-2886, 2013.

8. **Bo Yao**, Hani Hagras, Daniyal Al Ghazzawi, and Mohammed J. Alhaddad. " An Interval Type-2 Fuzzy Logic System for Human Silhouette Extraction in Dynamic Environments," The 2012 International Conference on Autonomous and Intelligent Systems, AIS'13, pp. 126-134. Springer Berlin Heidelberg, 2012.

9. **Bo Yao**, Hani Hagras, Mohammed Ghanbari, Mohammed J. Alhaddad, and Daniyal Alghazzawi, "A Type-2 Fuzzy Logic Approach for Detecting Human Related Events in Videos," UK Workshop on Computational Intelligence 2012, UKCI'12, vols. 1, pp. 1-6, 2012.

10. Xiao Yan, Dan Xu, and **Bo Yao**, "Large-scale Surveillance System based on Hybrid Cooperative Multi-Camera Tracking," Open Journal of Applied Sciences, vol. 3, no. 1b, pp. 79-84, 2012.

11. Xiao Yan, Junchuan Yang, and **Bo Yao**, "Design and DSP Optimization of Real-time Multi-Camera Tracking," Proceedings of Modelling, Identification and Control, Elsevier, MIC'12, vol. 3, pp. 589-594, 2012.

12. Junchuan Yang, Xiao Yan, and **Bo Yao**, "Character Feature Extraction Method based on Integrated Neural Network," Proceedings of Modelling, Identification and Control, Elsevier, MIC'12, vol. 3, pp.197-202, 2012.

13. Jingjie Jiang, Huaqing Tan and **Bo Yao**, "A Multi-platform Compatible and Open Software Architecture for Computer Numerical Control Systems," Advanced Materials Research, vols. 468-471, pp. 430-433, 2012.

Chapter 1: Introduction

The World Health Organization estimated that in 2050 there will be 1.91 billion people aged 65 years and over worldwide [WHO 2014]. Hence, recently there has been an increased interest in Ambient Assisted Living (AAL) technologies not only due to the increase of ageing population but also because of shortage of caregivers and the increasing costs of healthcare. Employing advanced machine vision based systems for behaviour and event detection as well as event summarization in AAL applications can help increase the level of care and decrease the associated costs. In addition, machine vision based systems can aid in detecting and summarizing important information which cannot be detected by any other sensor (like how frequently did the candidate walk, or whether the candidate is falling down, etc). However, the great expansion of deploying and utilizing video sensors can lead to massive amounts of unnecessary video data which can result in high costs because of data storage as well as because of the human resources spent on watching and/or manually extracting key video information. This problem is becoming increasingly obvious as the number of video cameras in use is estimated to be 100 million worldwide [Vos 2011], while in the United Kingdom, which owns the largest number of Closed-Circuit Television (CCTV) cameras in the world, the estimated number of in-use cameras is 5.9 million.

Conventional video systems based on human monitoring are highly labour-intensive since watching and analysing video content requires a higher level of concentrated attention. It has been reported that maintaining the necessary attention and reacting to rare events from multiple input video channels is a very challenging task, which is also extremely prone to error due to the degradation in the engagement level [Miller 1998]. Thus, there is a dramatically growing demand to develop real-time video detection and automatic linguistic summarization tools which are capable of autonomously detecting important events instantly and summarizing the interesting information from the massive raw video data in AAL applications in layman's terms. In order to automatically detect important events that need immediate attention, there is a need to analyse the real-time input data and provide valuable context information which cannot be extracted from other sensors. For example, an important application in elderly care within AAL environments ensures that the user drinks enough water throughout the day to avoid dehydration. The system also needs to send a warning message to social services nearby in case an elderly person falls and needs help so that proper actions can be taken instantly. Furthermore, in intelligent environments the electric appliances can be intelligently tuned and controlled according to the user's behaviour and activity to maximize their comfort and safety while minimizing the consumed energy.

Many AAL and healthcare applications based on behaviour and activity recognition have been reported in the literature. Single activity monitoring systems were proposed to analyse a single activity where in [Nambu 2008] a method was introduced to analyse the behaviour of watching TV for diagnosing health conditions. In [Wu 2008], researchers proposed an algorithm to analyse walking patterns in order to notify the elderly users in time so as to avoid the risk of falling down. However, a single activity analysis system is unable to recognize other important kinds of behaviour and it is not sufficient to create an effective AAL environment. In [Wan 2011], Wan et al. developed a behaviour recognition system to prevent the wandering behaviour of dementia patients and to notify the caretakers if deviation from predefined routes is detected. For the prevention of indoor stray, Lin et al. [Lin 2006] utilized Radio-frequency identification (RFID) sensors to detect if a dementia patient approaches an unsafe region in order to avoid potentially injurious situations. However, these kinds of locations and trajectory-based systems can only estimate the status of the subject via the position; they cannot recognize their actual behaviour and activity. Remote telecare systems can be constructed by using AAL based on activity recognition. For example, Barnes et al. [Barnes 1998] presented a low-cost solution to creating an intelligent telecare system by utilizing the infrastructure of British Telecom to assess the lifestyle feature data of the elderly. The system uses IR sensors, magnetic contacts and temperature sensors to collect the data of the temperature and the user's movement. An alarm notifies a remote telecare centre and the caregivers if abnormal behaviour is detected. However, the system is simple and is limited to only recognizing abnormal sleeping duration, uncomfortable environmental temperature, and fridge usage disarray. In another case, Hoey et al. [Hoey 2010] introduced a cognitive rehabilitation system using AAL technologies to help the elderly with dementia. Another cognitive orthotics system [Levinson 1997] analysed a model of the everyday activity plan according to multi-level events, and evaluated the patient's implementation of the plan for the purpose of cognitive orthotics. However, unlike our system, extendable recognition for complex behaviour and activity together with the summarization of the frequency, duration, timestamp and the user information has not been implemented in these conventional systems.

Most previous research on behaviour and activity recognition was based on 2D video data [Gowsikhaa 2014] or RFID sensors [Mehrjerdi 2014]. However, 2D video data based sensors are normally inadequate for capturing robust visual detailed features especially for those highly complex vision applications such as behaviour recognition. Hence, the use of 2D video data in real-world environments leads to relatively low accuracy due to the noise and uncertainties associated with sunshine, shadow, occlusion and colour similarity among other things. The use of RFID tags is intrusive and inconvenient as it requires a deployment of RFID tags on the human or objects. Dynamic models of behaviour characteristics can be constructed by utilizing statistics-based algorithms, for example Conditional Random Fields (CRF) [Ji 2010] and Hidden Markov Model (HMM) [Kim 2010]. However, the accuracy is not satisfactory. Dynamic Time Warping (DTW) is another classic algorithm [Reyes 2011] for behaviour recognition. However, DTW only returns exact values and thus it is inadequate for modelling the behaviour uncertainty and activity ambiguity.

Machine vision based behaviour recognition and summarization in real-world AAL is a very difficult task due to the high levels of encountered uncertainties caused by the large number of subjects, by behaviour ambiguity for different people, occlusion problems from other subjects (or non-human objects such as furniture) and environmental factors such as illumination strength, capture angle, shadow and reflection. To handle the high-levels of uncertainty associated with the real-world environments, Fuzzy Logic Systems (FLSs) have been employed. Various linguistic summarization methods based on Type-1 FLSs (T1FLSs) have been proposed; for example, [Anderson 2008], [Anderson 2009] employed T1FLSs for fall down detection. These type-1 fuzzy-based approaches perform well in predefined situations where the level of uncertainty is low. But these methods require multi-camera calibration, which is inconvenient and time-consuming.

In [Lara 2013], [Trivino 2008] T1FLSs were used to analyse the input data from wearable devices to recognize the behaviour and summarize human activity. However, such wearable devices are intrusive and can be uncomfortable and inconvenient as their deployment is invasive for the skin and muscles of the users. T1FLS was used in [Yao 2014], [Yaouanc 2012] to analyse the spatial and temporal features for efficient human behaviour recognition. In [Acampora 2015], a novel hierarchical neuro-fuzzy based algorithm was proposed for human behaviour analysis and recognition which enables both a quantitative and qualitative behavioral analysis that efficiently face the intrinsic people/objects tracking imprecision and provide context aware and semantic capabilities for better identifying a given activity. In [Acampora 2012], a novel algorithm based on combined neural networks and fuzzy systems was presented for the area of human behaviour understanding. This system [Acampora 2012] shows high level of scalability, robustness and tolerance for tracking imprecision and could represent a valid choice for improving the performance of traditional systems. In [Chang 2009] a fuzzy rule-based human activity recognition system for e-health was introduced and achieved an accuracy of about 90%. In [Medjahed 2009] human activities of daily living (ADL) recognition system for ubiquitous healthcare was proposed, which used hybrid sensors based on the fuzzy logic system, and the analysis result was robust. In [Almohammadi 2014], fuzzy logic was employed to recognize students' engagement degree so as to evaluate their performance in an online learning system. However, there are intra- and inter-subject variations in behavioural characteristics which cause high levels of uncertainty in the behaviour recognition. In [Yao 2014], [Almohammadi 2014], [Yao 2012], IT2FLS performed much better than T1FLS in human event detection and summarization.

In our research, we will present a robust behaviour recognition system for video linguistic summarization using the latest model of the 3D Kinect camera based on Interval Type-2 Fuzzy Logic Systems (IT2FLSs) optimised by Big Bang Big

Crunch (BB BC) algorithm [Erol 2006], [Kumbasar 2011] in order to obtain the parameters of the membership functions and rule base of the IT2FLS. We will present several real world experiments which were conducted for AAL related behaviour with various users. It will be shown that the proposed BB-BC IT2FLSs outperforms its Type-1 FLSs (T1FLSs) counterpart as well as other conventional non-fuzzy methods, and the performance improvement rises when the amount of subjects increases. It will also be shown that by utilizing the recognized output activity together with relevant event descriptions (such as video data, timestamp, location and user identification) detailed events are efficiently summarized and stored in our back-end SQL event database, which provides services including event searching, activity retrieval and high-definition video playback to the front-end user interfaces.

1.1 Aims, Significance, Novelty

1.1.1 Aims

In general, our research aims to contribute towards creating ambient assisted living environments which can intelligently understand the users by action and gesture recognition so that the needed healthcare services can be provided automatically and immediately to maximise user comfort and safety. In order to achieve this goal, in the fundamental first phase of this research, we will propose an effective and efficient system to extract human silhouette and to further detect particular human events. Then, we will propose a novel recognition algorithm based on fuzzy logic for human actions understanding. Specifically, various types of vision sensor such as 3D Kinect sensor (both Kinect v1 and Kinect v2) and CCTV (Closed Circuit TV) cameras were deployed in our smart ambient assisted living environment laboratory and in open outdoor/indoor public spaces to capture visual behavioural features of users' activity.

Through these vision sensors, we can analyse the users' behaviour and we are automatically able to detect particular events of interest.

1.1.2 Significance of this Research

For this research, we constructed the framework for ambient assisted living enabling the behaviour recognition and event summarization based on interval type-2 fuzzy logic system which is very good at handling the high-level of uncertainties in the environment being modelled. We constructed and deployed our system in real-world environments. According to the test feedback of the users, our method is robust in real-world usage. It is hoped that our framework has built an enabling system prototype of the ambient assisted living environments which helps reduce the stress caused by the dramatically increasing expenses being invested in healthcare and provide a reliable and affordable solution for smart healthcare and ambient assisted living. At the same time we are interested in automatically detecting human events from video streams for smart ambient environments, such as an intelligent living room and an intelligent classroom. The main challenge facing automatic event detection in videos is to detect human related events especially when the environment is complex as in the case of open public spaces. Hence, automatic human event detection is an essential prerequisite for many advanced video analysis systems.

1.1.3 Novelty

There are several challenges to the realization of this research's aim. These are human silhouette extraction, human behaviour recognition, event detection and summarisation, which our proposed methods have effectively addressed. The thesis provided possible solutions for behaviour recognition and event summarisation which can be applied in various applications including healthcare, surveillance, sport, entertainment and education, and others.

For human event detection, the first challenge is to segment (or outline) the human silhouette (human frame) from a video sequence. Since silhouettes in images and videos imply important information about foreground objects that can be utilized for high level analysis, silhouettes usually provide critical hints for image and video understanding and have a significant impact on the robustness of human related event detection. In our research, we will propose a robust method to eliminate the noise factors and extract the silhouette using interval type-2 fuzzy logic systems. Our method outperforms a similar method that is based on type-1 fuzzy logic systems by effectively reducing the misclassification. With a reliably extracted silhouette, various vision analysis algorithms and applications which focus on human subjects can be performed.

Based on the extracted silhouette, we here propose a recognition method for behaviour recognition using 2D vision sensor. The main advantage of 2D vision sensor (CCTV camera) is the hardware reliability and the convenience of its installation which is very beneficial for outdoor environments. However, the disadvantage is that the 2D video data captured form the camera is not robust enough to describe the behavioural features and it thus leads to relatively low recognition accuracy. We suggest a novel recognition framework based on interval type-2 fuzzy logic system and we tested our methods on the online dataset against various conventional methods including type-1 fuzzy logic system. Our method has the capability to acquire the parameters from the raw data and automatically tune the parameters in such way so as to adjust the fuzzy logic system to the changes of the environments. According to our results, our methods outperform the other conventional algorithms.

So far we have built an automatic learning and recognition framework which can find out the relevant parameters according to the input data so that more uncertainties and changes can be handled and accommodated. To further improve the performance and accuracy of behaviour recognition and event summarisation in relation to the main scenario, i.e., ambient assisted living, we chose a special sensor which particularly fits the requirements of constructing the ambient intelligence context awareness. We use Kinect v2 as our main sensor after taking into consideration its advantages and features, which are: 1) highly robust in sensing the visual and depth information in a short-middle range such that it can be applied perfectly in indoor environments and the nearby outdoor areas, 2) friendly low-cost (£130) but extremely reliable in indoor environments, and 3) detailed support documents and easy-to-use SDK providing the robust visual features which enable rapid research-and-development. However, one disadvantage is its request for an extra computer to run the analysis program. This causes a huge increase in the required budget for outdoor deployment since the price of special industrial computer which is robust against outdoor temperature and humidity is high (around £800-£1000). The other disadvantage is the inconvenience of its deployment caused by its shape and its cables. These problems can be solved by improving the hardware platform of the sensor. This can be done by utilizing an embedded system rather than a computer.

Based on the Kinect v2, we developed a robust analysis system for behaviour recognition and event summarisation. Our system firstly captures the sensing data from the Kinect v2, and then analyse the input data by our proposed IT2FLS-based method for robust 3D behaviour recognition to understand the current behavioural status of the human subject. After that, based on the results of the behaviour recognition, to condense the massive raw information into key events, the event summarisation will be performed by combining all the related details key information (e.g. subject identification, subject number, behaviour category, event time stamp, event video data, etc.) regarding the detected behaviour and summarise them into an event which will then be sent and stored into the back-end event database so that event retrieval and playback can later be performed by the users using the front-end GUI system.

Due to the limitation of the field-of-view and sensing distance of the hardware platform of the Kinect, according to our experiments, we found out that the reliability of the sensing data 3D skeleton will degrade if there are crowded multiple users since some of the users will be around the boundary of the view or there would be serious occlusion problem. And in our research we only focus on behaviour recognition and event summarisation, in our current system, we don't recognise or detect who the people are. And this module of person identification will be developed in our future work.

To the best of the author's knowledge, this work is the first in the literature to apply a type-2 fuzzy logic system to visual-based humans' behaviour recognition and event summarisation and to construct a practical solution and system prototype for the application of ambient assisted living.

1.2 Research Objective

The general research objective of my research is to construct an ambient assisted living environment by proposing and using novel computer vision techniques, type-2 fuzzy logic systems, and optimisation algorithms. For the entire vision-based AAL system, the basic step is to segment the human silhouette from the original video sequence so as to extract significant information (such as area, position, shape, contour, etc.) of the human target. After that, our proposed IT2FLS-based recognition system will be carried out to understand the behaviour of the human subject. And then, to further improve the robustness and performance of our AAL system, we will present 3D sensor-based event summarisation system for robust event detection in indoor AAL environment. Our system is designed to detect the important AAL behaviours and to summarise the target events happened in the AAL environments. To further guarantee the adaptiveness, the parameters of our system is designed to be automatically optimised by our system.

1.3 The Structure of the Thesis

The rest of the thesis is organized as follows. In Chapter 2, we will provide an overview of the relevant techniques and aspects employed in the thesis. Specifically, we will provide an overview on type-1 and type-2 fuzzy logic systems and we will give an outline of human behaviour recognition and event summarisation and of the various approaches that were taken so as to solve the problem. Chapter 3 presents our proposed type-2 FLSs based method for human silhouette extraction which is the first step in event detection and summarisation. Chapter 4 introduces the proposed BB-BC based IT2FLS for the behaviour recognition on 2D video data. In Chapter 5, we proposed BB-BC based IT2FLS for the event detection and summarisation using 3D sensor in real-world ambient assisted living environments. Finally, both our conclusions and our plan for future work are presented in Chapter 6.

1.4 Discussion

In this chapter, we introduced behaviour recognition and event summarisation in ambient assisted living. Behaviour recognition enables the computational intelligence to understand what the subject (human user) is doing and it classifies the subject's current action/behaviour/gesture into the learned categories of the knowledge base. After that and based on the results of behaviour recognition and other relevant information such as time stamp, location, user's information among others, event summarisation can be performed to abstract the detected key information into events based on which smart services, such as healthcare or security surveillance, can be provided automatically and instantly to the users so as to maximize user comfort and safety.

For our research, we also compared our proposed system with the recent commercial smart vision system so as to demonstrate the potentials and advantages of our proposed system. Generally commercial cameras consist of several major modules: hardware platform, management software, and analysis intelligence. During the past decades, huge improvements were conducted on the hardware platform of the vision system. For example, there were changes from local analogue signal-based camera to network IP signal-based camera, from PC-based smart vision system to embedded integrated smart vision system, from standard-definition (640×480) resolution to high-definition (1920×1080) resolution and so on. Meanwhile, the software (management, storage, compression) was simply adapted to the development of the hardware. With the development of the hardware, the analysis intelligence has become increasingly feasible and available for the commercial vision system in the recent years. However, the existing commercial vision systems barely focus on "Tracking" to obtain the bounding box (area and position) and trajectory (movement history) of the target. Several basic event detections based on the trajectory can be performed such as when the target crosses a border line, enters the dangerous area, approaches the door/car, and so on. Important events cannot be distinguished effectively. No high-level analysis is conducted "inside the bounding box" to achieve a detailed and deep-level recognition of a human's behavioural features. For example, a person being at a specific location of a living room can be detected but it remains an unsolved problem to distinguish whether the subject is falling down on the floor or is sitting safely and comfortably on the sofa since the recent commercial vision systems are trajectory-based and they are unable to perform detailed behaviour recognition. This problem generates the gap and blank between the vision (camera) system and real-world applications. In our research, most of the famous international companies (such as HikVision, HoneyWell, Axis Communication, Bosch, Samsung, Mobotix, Sony, etc.) which manufacture cameras have been investigated so as to prove that this problem exists. In order to fill this gap and further address this problem, a higher-level analysis system is developed, which performs a detailed and deep-level recognition of the human's visual cues and behavioural features as well as a trajectory analysis. As it can be seen in our experiment results of the remaining chapters, high-level important/interested events can be robustly detected and efficiently summarised in intelligent ambient assisted living environments in our proposed system. In this way, an advanced smart vision system which is capable for both large-scale monitoring and detail-level features analysis is successfully constructed. In a similar manner, an advanced system can be developed for the applications including healthcare, automation and security in intelligent ambient assisted living environments.

To sum up, our research aims to develop a type-2 fuzzy logic based analysis system that enables human behaviour recognition in machine vision systems in order to fill the gap in the current commercial vision systems and construct high-level applications such as healthcare, education, automation and surveillance in ScaleUp intelligent environments, which is also the major development trend of machine vision in both academia and industry.

Chapter 2: Overview of the Relevant Techniques and Aspects Employed in the Thesis

2.1 Overview on Fuzzy Logic Systems

Logic-based algorithms are a computational model, which is able to emulate nature in problem solving. Logic-based methods provide a means of reasoning that allows the system to receive input values and use the reasoning method to analyse them in order to provide the corresponding output values. Overall, there are boolean logic and fuzzy logic methods.

2.1.1 Boolean Logic Methodology

One major feature of the boolean logic model is that the inputs are discrete. Specifically, the input value of boolean logic can either be 0 or 1, and the relation between inputs are modelled as mixtures of the logical operator '*AND*', '*OR*' and '*NOT*' based on which, the generated rules are computed in the reasoning procedure of the boolean logic system. Basically, the generated rules come from a truth table which consists of lookup values mapping all the possible combinations of the inputs to their corresponding outputs where the output membership value is either 0 or 1 [Kassem 2012].

Besides boolean values, the boolean logic also uses crisp sets. According to [Mendel 1995], *a* crisp set A in a universe of discourse U can be represented by finding the elements x ϵ A which is performed by finding a condition by which x ϵ A such that $A = \{x \mid x \text{ meets the condition}\}$. This means that for any element x who belongs to the set A, x has a membership degree of 1, otherwise if x does not belong to A, then x has membership degree 0 [Mendel 1995].

Figure 2.1 shows an example of a crisp set which models tips (in a restaurant). It shows that if the tip amount is less than or equal to £3, then this tip amount is cheap and belongs to the crisp set of cheap tip whose membership degree is 1. Otherwise this tip amount does not belong to the crisp set and the membership degree by which it belongs to the set "Cheap" is 0.

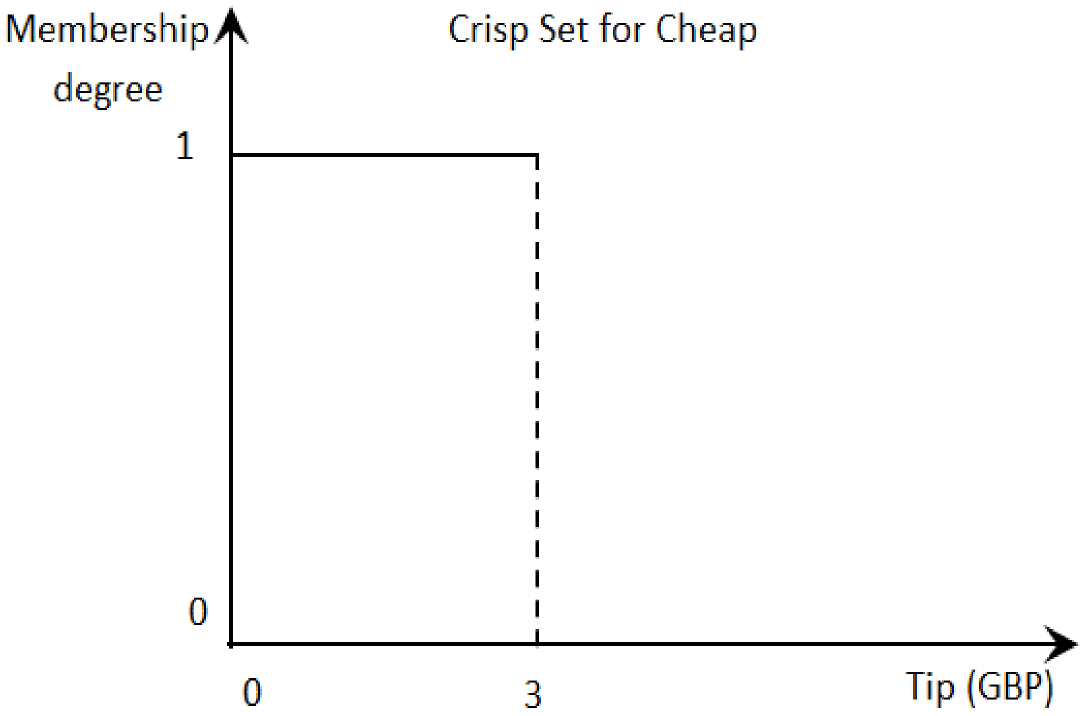


Figure 2.1: Example of a Crisp Set

The main feature of this type of model is that it is simple to understand and easy to implement; however, it is incapable of handling a variety of situations [Aldridge 2009]. This is because it is very difficult to define the turning point between the concepts "cheap" and "generous". There are differences in people's opinions and some people may define £2 as the turning point while others may define £5 as the turning point. Due to this problem, this kind of system cannot handle the uncertainty created or the imprecise inputs which are very common in real-world scenarios. To solve this problem and better handle the uncertainty, fuzzy logic was proposed as well as widely used.

2.1.2 Fuzzy Logic Methodology

Fuzzy logic is an analysis methodology which is closer to the way a human solves a problem rather than the conventional logic based methods. Fuzzy logic is capable of providing a robust way to model the approximate, inexact nature which commonly exists in the real-world problems, and it represents the system behaviour via a rule base where each rule consists of antecedents and consequents [Kassem 2012].

One major analysis feature of fuzzy logic is that it converts linguistic control procedures into automatic control procedures by using expert knowledge. A fuzzy logic system especially performs better than traditional control methodologies in the tasks where the data are too complex to be processed by traditional analysis methods or when there are high-levels of uncertainties in the input sources of data causing huge inexact blurred ambiguity.

A difference between fuzzy logic and boolean logic is that boolean logic uses a two-state representation while fuzzy logic uses membership degrees which range between 0 to 1 [Kassem 2012]. Thus, a single input of fuzzy logic can have multiple membership degrees belonging to more than one set.

2.1.3 Fuzzy Sets

Fuzzy set was introduced by [Zadeh 1965] and it has been frequently extended by researchers in the recent decades. Owing to the advantages of fuzzy set, we can impose different structures on the membership space. There are several types of available fuzzy sets in the literature. Here we mainly focus on type-1 fuzzy set and type-2 fuzzy set, which are important in recognition and classification applications.

Basically, a fuzzy set is a class of objects with continuous membership degrees [Zadeh 1965]. Thus the membership function assigned to each element *X* in a universe of discussion with membership degree ranging between *0* and *1* and the non-membership degree equals one minus the membership degree. This membership degree combines the evidence for *X* and the evidence against *X* [Dubois 2000]. The definition that fuzzy sets are based on is:

Definition 2.1 [Zadeh 1965] Let *X* be a space of points (object), with a generic element of *X* denoted by *x*. Therefore, $X = \{x\}$. A fuzzy set (class) A in X is characterised by a membership (characteristic) function $f_A(x)$ which associates with each point in X a real number in the closed interval [0, 1] with the value of $f_A(x)$ at *x* representing the "grade of membership" of *x* in *A*. Specifically, a fuzzy set in a classical set *X* is defined as follows:

$$A = \{(x, \mu_A(x)) | x \epsilon X\} \qquad (2.1)$$

The closer the degree of $f_A(x)$ to unity is, the higher the membership degree of *x* in *A is*. When *A* is a set in the ordinary sense of the term, its membership function can take on only two values 0 and 1, with $f_A(x)$ or 0 according as x does or does not belong to *A*. When it is necessary to differentiate between such sets and fuzzy sets, the sets with a two-state characteristic function will be referred to as ordinary sets or simply sets [Zadeh 1965].

Definition 2.2 [Zadeh 1965] The support of a fuzzy set *A*, *S(A)*, is the crisp set of all $x \epsilon X$ such that $\mu_A(x) > 0$

Definition 2.3 [Zadeh 1965] The (crisp) set of elements that belongs to the fuzzy set *A* at least to the degree α is called the α-level set:

$$A_\alpha = \{x \epsilon X | \mu_A(x) > \alpha\} \qquad (2.2)$$

$A_\alpha = \{x \epsilon X | \mu_A(x) > \alpha\}$ is called "strong level α – level set" or "strong α cut".

Definition 2.4 [Zadeh 1965] A fuzzy set *A* is convex if

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_1)\}, x_1, x_2 \epsilon X, \lambda \epsilon [0,1] \quad (2.3)$$

Alternatively, a fuzzy set is convex if all α-level sets are convex.

Definition 2.5 [Zadeh 1965] For a finite fuzzy set *A*, the cardinality |*A*| is defined as:

$$|A| = \sum_{x \epsilon X} \mu_A(x) \quad (2.4)$$

$||A|| = \frac{|A|}{|X|}$ is called the relative cardinality of *A*. The relative cardinality of a fuzzy set has to be in the same universe when making a comparison of fuzzy sets by their relative cardinality.

Figure 2.2 displays an example of a type-1 fuzzy set. Figure 2.3 shows the support of a fuzzy set where the support of a fuzzy set is the set of all points x in X such that $\mu_A(x) > 0$. Figure 2.4 shows an example of a singleton fuzzy set, where a singleton is a fuzzy set whose support is a single point in where $\mu_A = 1.0$ [Kassem 2012].

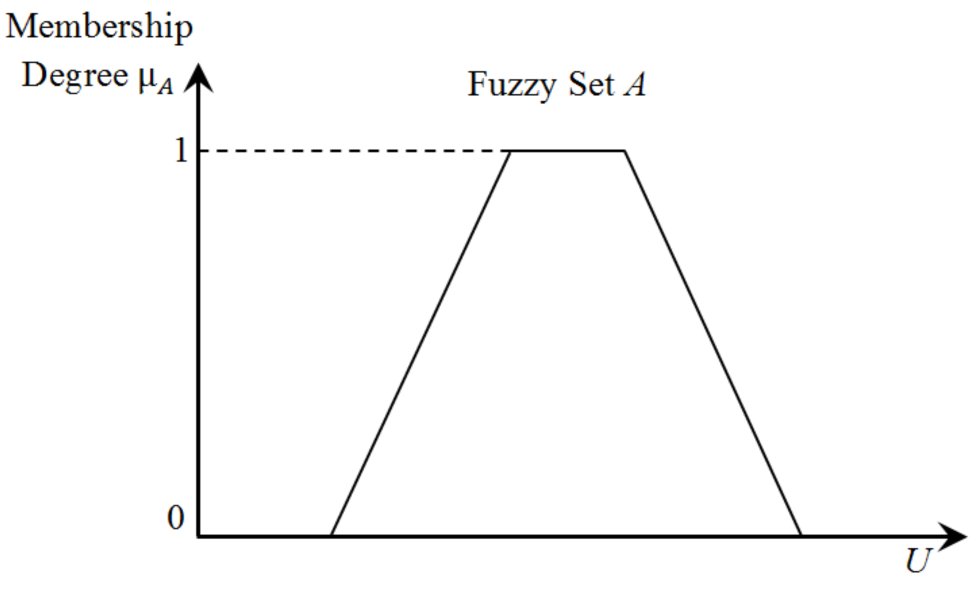


Figure 2.2: Example of a type-1 fuzzy set.

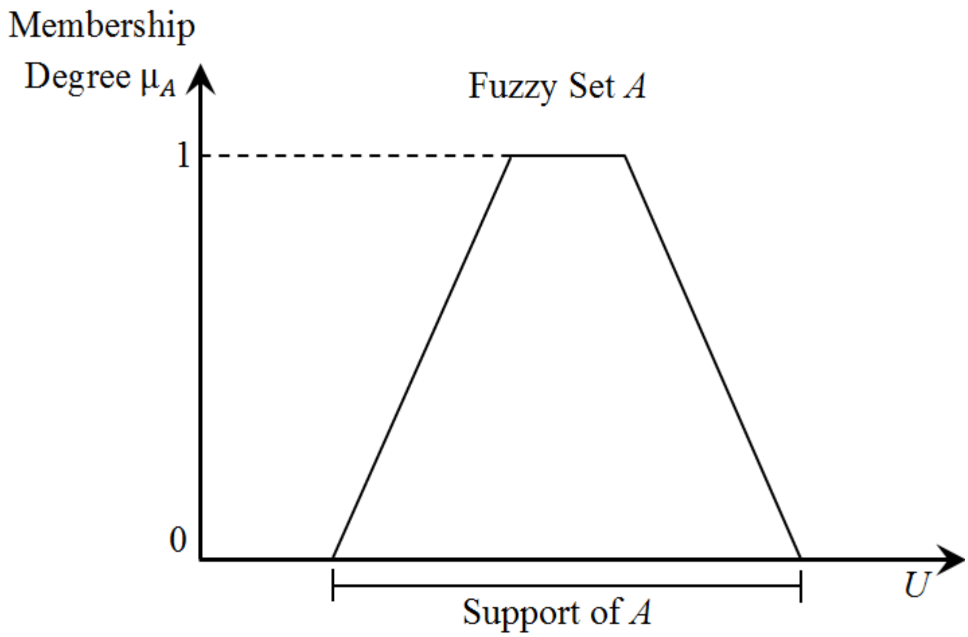


Figure 2.3: Support of a Fuzzy Set.

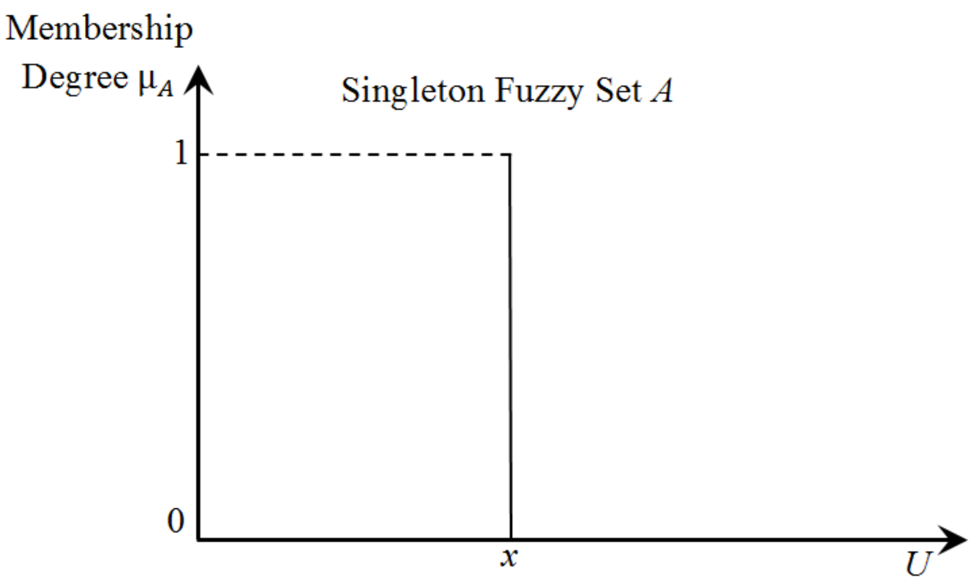


Figure 2.4: Example of a Singleton Fuzzy Set.

2.1.4 Operations with Fuzzy Sets

The analysis procedure of fuzzy sets allows the calculating of the sets in a general way. In order to conduct operations on fuzzy sets such as union or intersection, set operations are applied such as t-norm (for intersection) and t-conorm (for union) on the fuzzy sets' membership functions, and the most commonly used t-norms and t-conorms in fuzzy logic engineering applications are the product or minimum t-norm and the maximum t-conorm [Mendel 1995]. Namely, *A*, *B* are fuzzy sets with membership functions μ_A, μ_B , respectively. The following equations show the intersection and the union between *A* and *B*, respectively, where * is the chosen t-norm and $\oplus$ is the chosen t-conorm [Mendel 2000].

$$A = \int_v A(v)/v \tag{2.5}$$

$$B = \int_w B(w)/w \tag{2.6}$$

$$A \cap B = \mu_A(v) * \mu_B(w) \tag{2.7}$$

$$A \cup B = \mu_A(v) \oplus \mu_B(w) \tag{2.8}$$

Then the complement $\overline{A}$, union $A \cup B$ and intersection $A \cap B$ are also fuzzy sets, and their membership functions are defined for $x \epsilon X$ by [Klir 2001].

$$\mu_{\overline{A}}(x) = 1 - \mu_A(x) \tag{2.9}$$

$$\mu_{\mathrm{A \cup B}}(x) = \max\big(\mu_A(x),\ \mu_B(x)\big) \tag{2.10}$$

$$\mu_{\mathrm{A \cap B}}(x) = \min\big(\mu_A(x),\ \mu_B(x)\big) \tag{2.11}$$

Illustrative examples of union, intersection, and complement of fuzzy set are demonstrated in Figure 2.5, 2.6, 2.7, 2.8 [Kassem 2012].

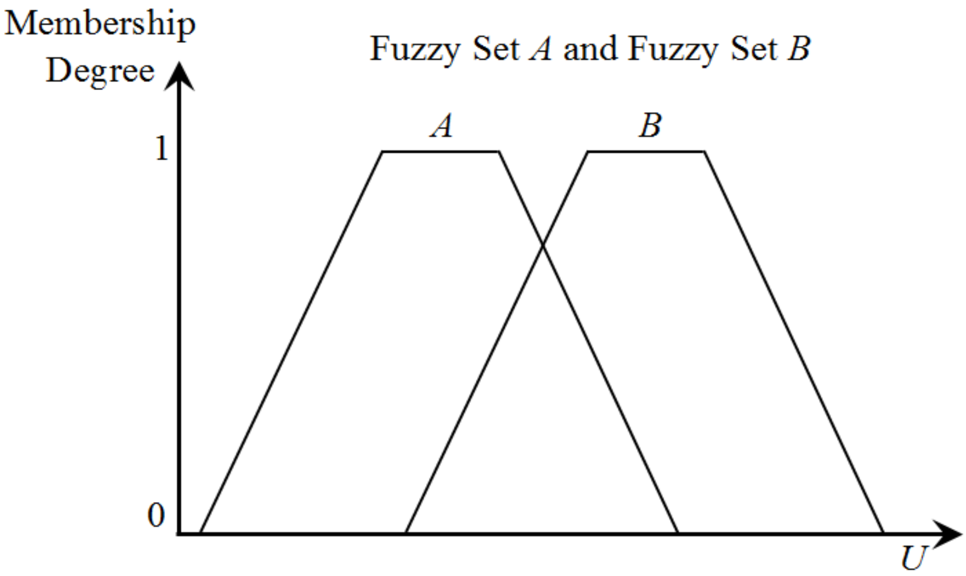


Figure 2.5: Two Fuzzy Sets A and B.

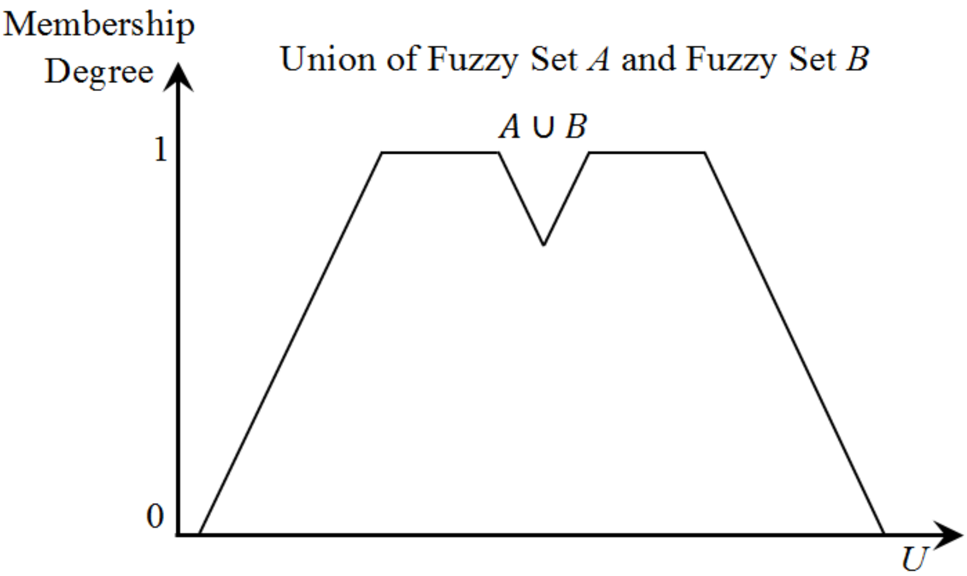


Figure 2.6: The Union of two A and B.

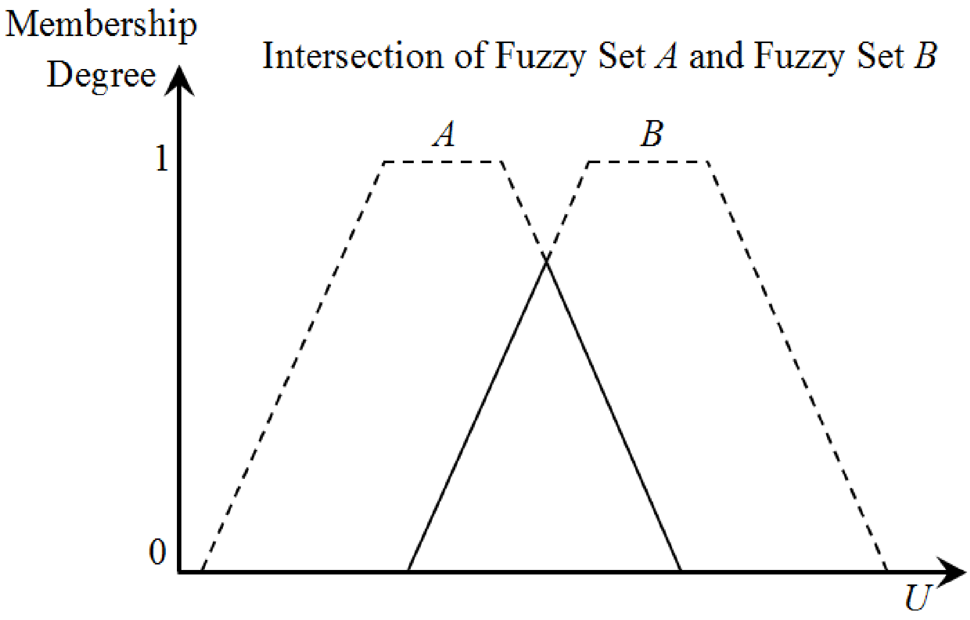


Figure 2.7: The Intersection of two Fuzzy Sets A and B.

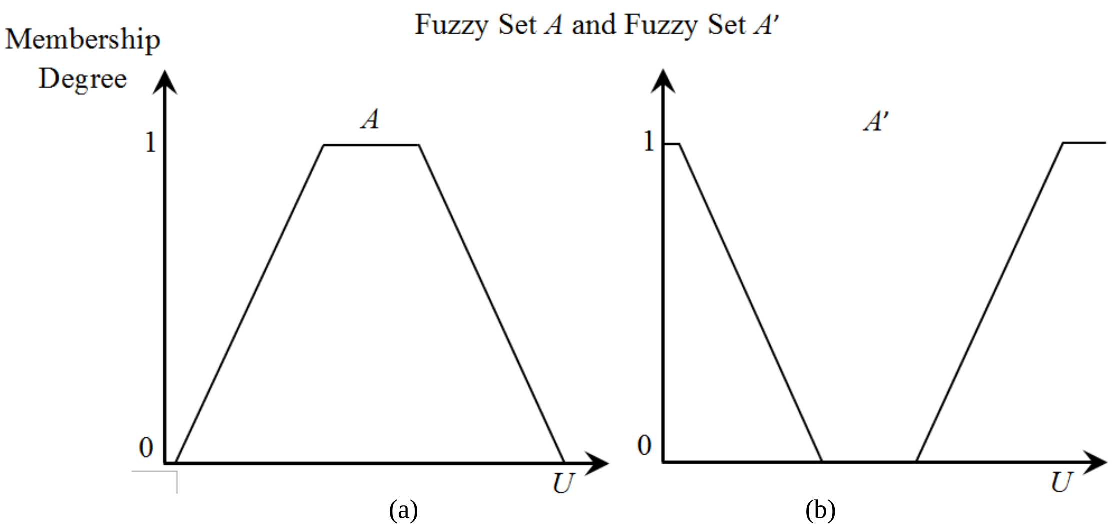


Figure 2.8: a) Fuzzy Set A. b) The Complement of a Fuzzy Set A.

As it can be seen from the above figures, fuzzy sets can be used for the manipulation of vague imprecise concepts; in particular a fuzzy set can be employed to represent linguistic variables [Kassem 2012]. Moreover, a linguistic variable is a term whose values are fuzzy or whose values are defined in other linguistic terms [Lee 1990].

2.1.5 Linguistic Variables

Figure 2.9 shows an example that explains the concept of fuzzy variables. The linguistic variable in fuzzy logic system is a concept which represents a certain input. For each linguistic variable, there is a corresponding set of terms denoting the different fuzzy sets. Each of the linguistic terms has a corresponding fuzzy set representing linguistic concepts such as “very small”, “small”, and “medium” and so on and so forth. In linguistic variables there subsist linguistic hedges such as “very”, “more” or “less” and “extremely” which stress the linguistic modifiers. (v, T, X, G, m) such that linguistic variable may be characterised by a quintuple where, [Klir 2001]:

v - is the name of the variable.

T - denotes the set of linguistic terms of v; that is the set of the name of linguistic value of x, with each value being a fuzzy variable denoted generically by x and ranging across a universe of discourse X which refers to a base variable whose values range over a universal set X.

G - is a syntactic rule for generating linguistic terms, and m is a semantic rule that is assigned to each linguistic term, t ϵ T its meaning, $m(t)$, which is a fuzzy set on X (m: $T \rightarrow f(X)$).

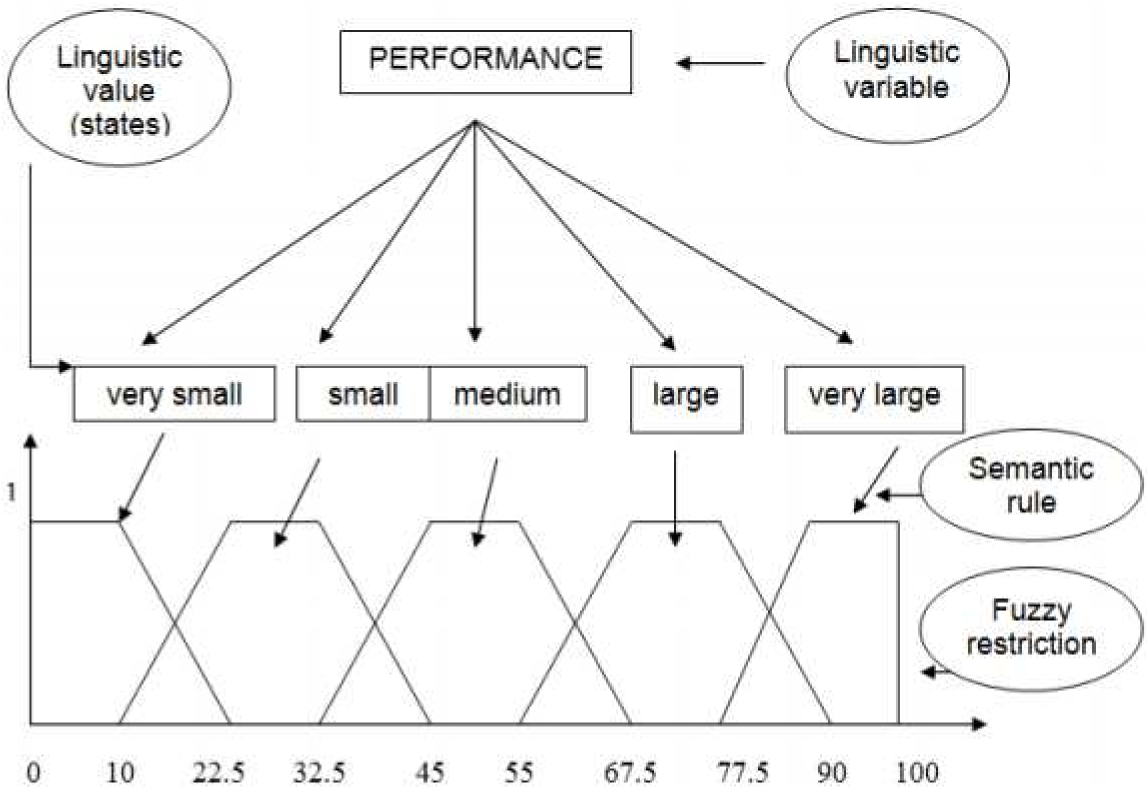


Figure 2.9: Example of Linguistic Variables [Klir 2001]

An example of a linguistic variable performance, which is taken from [Klir 2001], is given in Figure 2.9

For further clarification, we use here an example of the temperature of the weather to explain the linguistic variable and membership functions. As shown in Figure 2.10 for example, if *temperature* is a linguistic variable, then its term set *T* (*temperature*) could be {*cool*, *warm*, *hot*}. Each term in this set is characterized by a fuzzy set in a universe of discourse *X* = [*0, 100*]. "*cool*" expresses that "*the temperature is below 10 degrees*", "*warm*" expresses that "*the temperature is below 20 degrees*" and "*hot*" expresses that "*the temperature is below 30 degrees*". As it can be seen in Figure 2.10, the membership functions and the linguistic variable's terms are displayed.

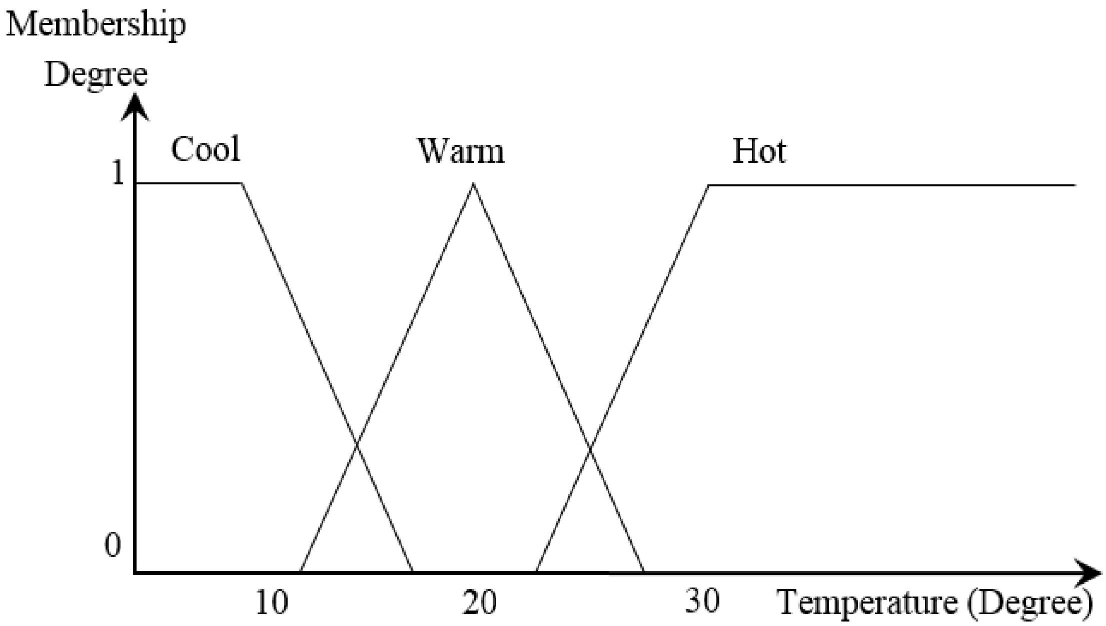


Figure 2.10: Linguistic Variable Temperature's Membership functions.

There are plenty of types of membership functions that can be used including triangular, trapezoidal, Gaussian function and others [Cirstea 2002] [Kassem 2012].

- Triangular: the triangular membership function (shown in Figure 2.11 and represented by Equation (2.12)) depend on three parameters *a*, *b* and *c*; where *a* is the starting point, *b* is the vertex and *c* is the end point of the triangle. The triangular function can be given by:

$$f(x\colon a,b,c) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & x > c \end{cases} \tag{2.12}$$

- Trapezoidal: the trapezoidal membership function (shown in Figure 2.12 and represented in Equation (2.13)) depends on four parameters, a, b, c and d. a

and d are the start and end points, b and c are the two points in between. The trapezoidal function is given by:

$$f(x\colon a,b,c) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & x > d \end{cases} \tag{2.13}$$

- Gaussian: the Gaussian membership function (shown in Figure 2.13 and represented in Equation (2.14)) depends on two parameters and m; where σ is the standard deviation and m is the mean. The Gaussian function is given by:

$$f(x\colon \sigma, m) = exp\left[\frac{-(x-m)^2}{2\sigma^2}\right] \tag{2.14}$$

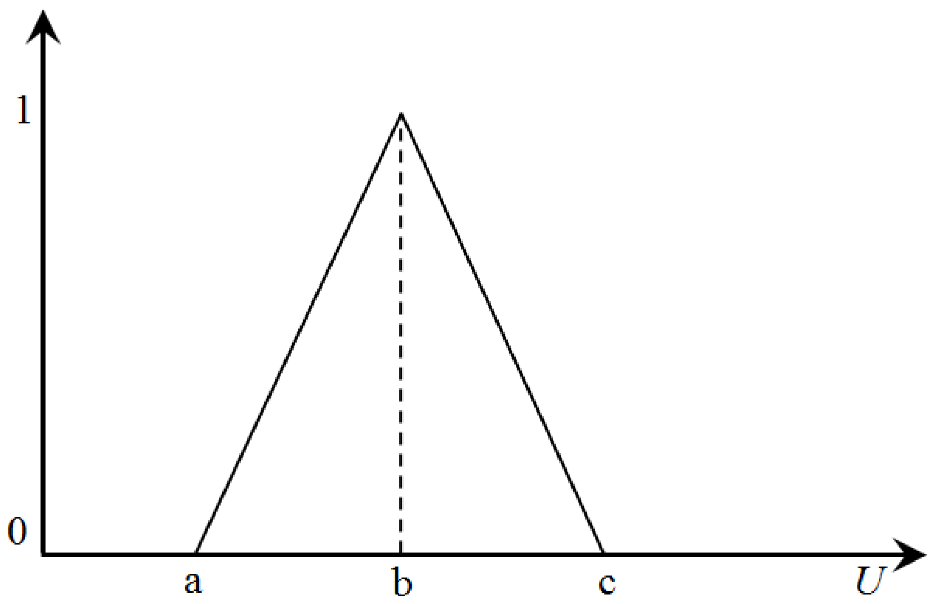


Figure 2.11: Triangular Membership Function.

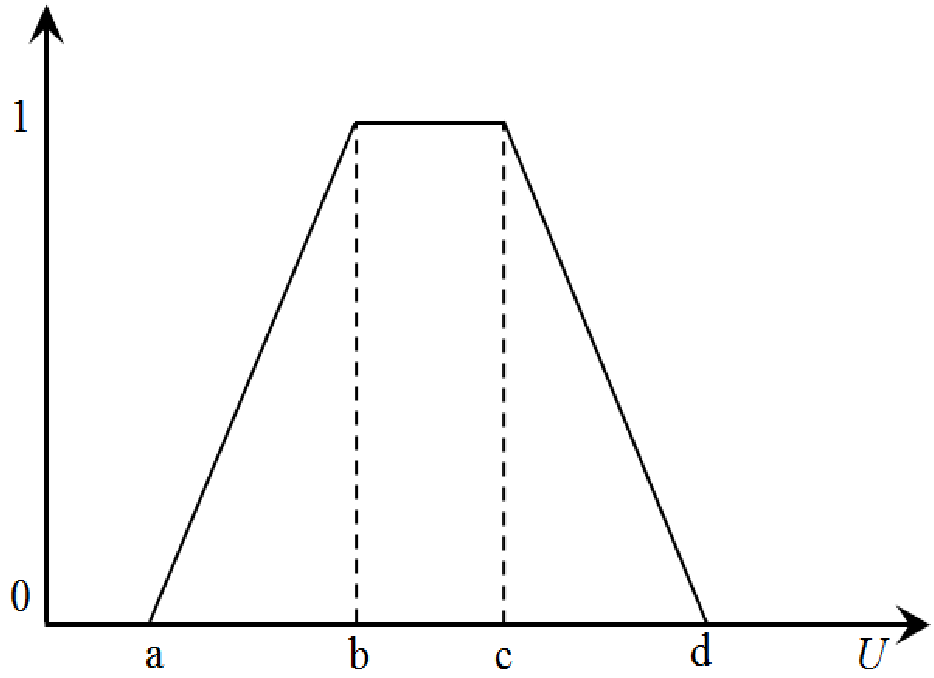


Figure 2.12: Trapezoidal Membership Function.

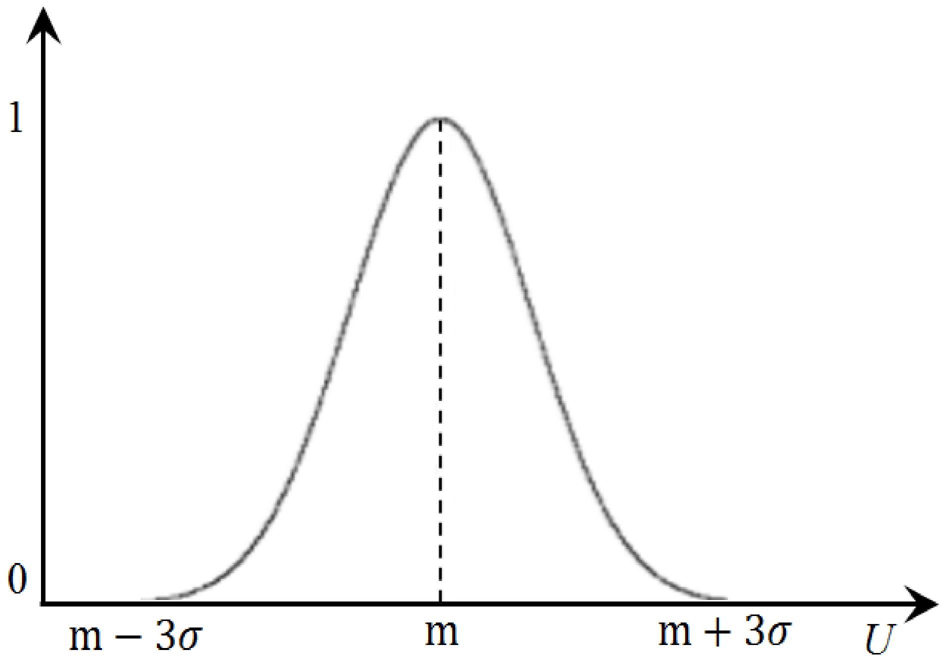


Figure 2.13: Gaussian Membership Function.

So far we have discussed how the fuzzy logic systems convert a crisp number into a linguistic variable and assign a membership value to it. In what follows, we will discuss the main modules of a type-1 fuzzy logic system with more details.

2.1.6 Type-1 Fuzzy Logic Systems

A type-1 fuzzy logic system maps crisp inputs into crisp outputs. It consists of four major components. These are the fuzzifier, rule base, inference engine, and defuzzifier as shown in Figure 2.14.

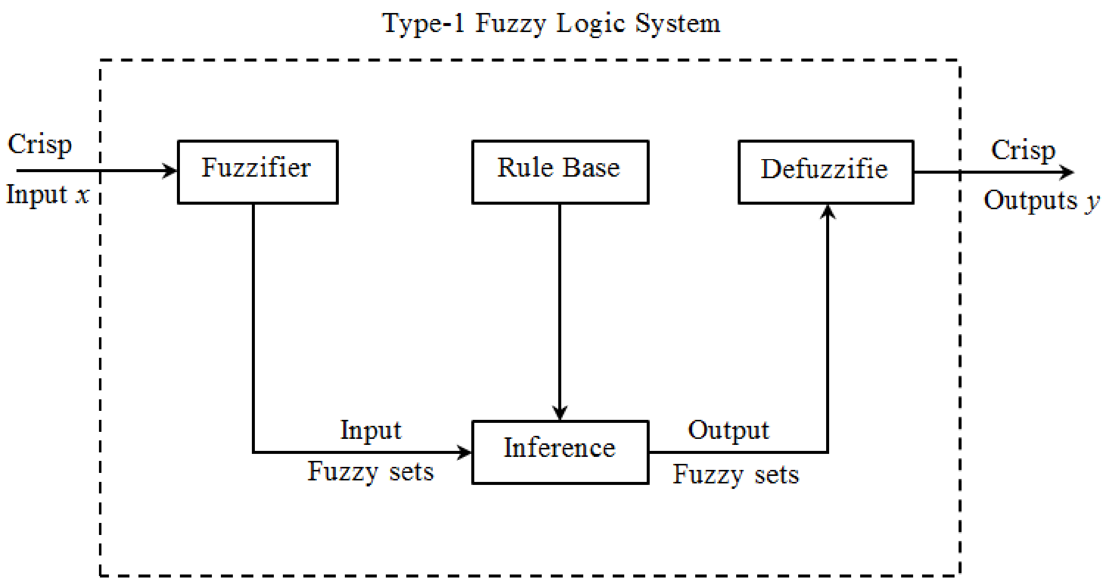


Figure 2.14: Type-1 fuzzy logic system

2.1.6.1 Fuzzification

The fuzzifier maps input crisp value into fuzzy sets. This procedure is done by mapping each input to its corresponding set, and accordingly finding the membership value by which this input belongs to the sets [Lee 1990].

There is more than one method of fuzzification. In our research we mainly focus on singleton fuzzifier. Singleton fuzzification unit converts a crisp value into a fuzzy singleton within a certain universe of discourse and a fuzzy singleton is a precise value; hence no fuzziness is introduced by the fuzzification in this case [Kassem 2012]. Singleton fuzzification expresses an input x_0 as a fuzzy set A with the membership function $\mu_A(x)$ equal to zero unless at the point x_0 at which $\mu_A(x_0)$ equals one [Mendel 1995].

2.1.6.2 Rule Base

The rule base is composed of the rules representing the routine and paradigm that the system behaves. In each rule, there are antecedents and consequents in an If-Then statement. Antecedents are represented by the fuzzy sets of input linguistic variables of the fuzzy logic system while consequents are represented by the fuzzy sets of output linguistic variables. In order to obtain the result outputs, the system needs to map the input crisp value into fuzzy sets; therefore logic rules need to be activated which are in terms of linguistic variables which are associated with the logic rules [Mendel 1995]. Rules may be given by experts or can be extracted from the raw numerical data [Mendel 1995].

2.1.6.3 Inference Engine

The inference engine of fuzzy logic system maps fuzzy sets into fuzzy sets [Mendel 1995].

Equation (2.15) shows an example [Kassem 2012] of a rule from the fuzzy rule base:

$$R^l: IF\ u_1 \text{ is } F_1^l \text{ and } u_2 \text{is } F_2^l \ldots \text{and } u_{\mathrm{p}} \text{ is } F_p^l \text{ THEN } v\ is\ G^l \tag{2.15}$$

where *l=1,2,…M*, *M* is the number of rules in the rules base, *p* is the number of inputs. $F_1^l, F_2^l,\ldots, F_p^l$ are fuzzy sets in U_1, $U_2,\ldots, U_{\mathrm{p}}$ and G^l are fuzzy sets in *V*. The inference engine uses these IF-THEN rules to map the input sets in $U = U_1 \times$

$U_2 \times ... \times U_p$ to output sets in *V*. Each rule in the rule base can be interpreted as a fuzzy implication. Specifically the 'THEN' operator is modelled by using the fuzzy implication $\rightarrow$ [Kassem 2012]. Let $F_1^l \times F_2^l \times ... \times F_p^l$ be *A*, and G^l be *B*, the rule shown in Equation (2.15) is interpreted by the inference engine as $A \rightarrow B$ [Mendel 1995].

The mapping is done from $\mu_A(u)$ to $\mu_B(v)$, where $u \in U$ and $v \in V$. *u* and *v* are linguistic variables, their numerical values are *x* and *y* where $x \in U$, $y \in V$ [Kassem 2012], and by taking *x* and *y* into consideration the interpretation performed by the inference engine can be written as in Equation (2.16).

$$\mu_{R^l}(x, y) = \mu_{A \rightarrow B}(x, y) \tag{2.16}$$

To compute the firing strength $f^l(x)$ of the l^{th} rule R^l, the calculation shown in Equation (2.17) is conducted, where * is the chosen t-norm.

$$f^l(x) = \mu_{x_1}(x_1) * \mu_{x_2}(x_2) * ... * \mu_{x_p}(x_p) \tag{2.17}$$

After the calculation of the firing strength $f^l(x)$ for each rule R^l we can determine the output fuzzy set B^l. The final output fuzzy set *B* is determined by combining the output fuzzy set for each rule using a t-conorm operator [Kassem 2012].

2.1.6.4 Defuzzification

The defuzzifier map output sets into a crisp number. There are several methods for defuzzification, including [Mendel 1995]:

- Maximum Defuzzifier: this defuzzifier chooses as its crisp output the value *y* from the set *B* for which $\mu_B(y)$ is the maximum.

- Mean of Maxima Defuzzifier: this defuzzifier starts by determining the values of y from the set for which $\mu_B(y)$ is the maximum. It then computes the mean of these values as its crisp output.

- Centroid Defuzzifier: This defuzzifier uses Equation (2.18) in order to determine the centre of gravity (centroid) $\overline{y}$ of the output fuzzy set B and uses this value as the crisp output. The summation from $i = 1$ till I denotes the support of the fuzzy set B.

$$\overline{y} = \frac{[\sum_{i=1}^{I} \mathrm{y}_i \mu_B(\mathrm{y}_i)]}{[\sum_{l=1}^{l} \mu_B(\mathrm{y}_i)]} \tag{2.18}$$

- Centre of sets Defuzzifier: This defuzzifier determines the centroid y^l for the fuzzy sets B^l (which is associated with the rule R^l). Then the crisp output y_h is calculated using Equation (2.19)

$$\overline{y} = \frac{\left[\sum_{l=1}^{M} y^l \mu_{B^l}(y^l)\right]}{\left[\sum_{l=1}^{M} \mu_{B^l}(y^l)\right]} \tag{2.19}$$

$\mu_{B^l}(y^l)$ represents the firing strength of the l^{th} rule.

2.1.7 Interval Type-2 Fuzzy Logic Systems

A fuzzy logic system is capable of handling the various sources of uncertainties associated with the environments being modelled. Hence there is a need for a system that does not need to be frequently redesigned or tuned in order to accommodate for the changes resulting from the high-levels of uncertainties [Hagras 2007].

Unfortunately, for human behaviour and event detection, there are plenty of sources of uncertainties in an environment. Behaviour features of different subjects which are representative of the same action classes have a wide variance. In addition, the behaviour of a given subject while conducting multiple instances of the same action category is not unique. Thus there are intra- and inter- subject variations in behavioural characteristics which cause uncertainty in the behaviour recognition problem. Besides this intra- and inter- subject uncertainty, there are also plenty of noise factors, such as the occlusion problem caused by the crowdedness of multiple subjects, and the objects in the environment hiding parts of the visual behaviour features from the camera.

The main procedure of Fuzzy Logic Systems (FLS) which enables FLS to handle a high-level of uncertainties is the conversion from crisp values into linguistic variables based on fuzzy sets [Kassem 2012]. At this point the main objective is that the uncertainties associated with environment are modelled into uncertainties by using the fuzzy set membership functions. Most of the fuzzy logic systems which have been used and reported so far were based on the conventional type-1 fuzzy logic systems, which use type-1 fuzzy sets. However, type-1 fuzzy sets cannot fully handle or accommodate the linguistic and numerical uncertainties because they use precise type-1 fuzzy sets; thus once the type-1 membership ship functions have been chosen all uncertainty disappears because type-1 membership functions are precise [Mendel 2001]. In order to solve this problem and handle the uncertainties, type-2 fuzzy sets are used in type-2 fuzzy logic system.

2.1.7.1 Type-2 Fuzzy Sets

According to [Mendel 2002], the concept of a type-2 fuzzy set was introduced by [Zadeh 1975] as an extension of the concept of an ordinary fuzzy set (type-1 fuzzy set). A type-2 fuzzy set is a set in which we also have uncertainty about the membership function, i.e., a type-2 fuzzy set is characterised by a fuzzy membership function whose grade for each element is a fuzzy set [0, 1].

The membership functions of type-2 fuzzy sets are three dimensional and include a footprint of uncertainty; it is the new third-dimension of type-2 fuzzy sets and the footprint of uncertainty that provide additional degrees of freedom that make it possible to directly model and handle uncertainties [Kassem 2012].

Generally, type-2 fuzzy sets can take any shape or value in the third dimension; this is called general type-2 fuzzy sets. In the interval type-2 fuzzy sets all the third dimension values are equal to one [Mendel 2001]. The use of interval type-2 FLS helps simplify the computation (as opposed to the general type-2 FLS) [Mendel 2001]. Figure 2.15 shows an example of an interval type-2 fuzzy set [Kassem 2012].

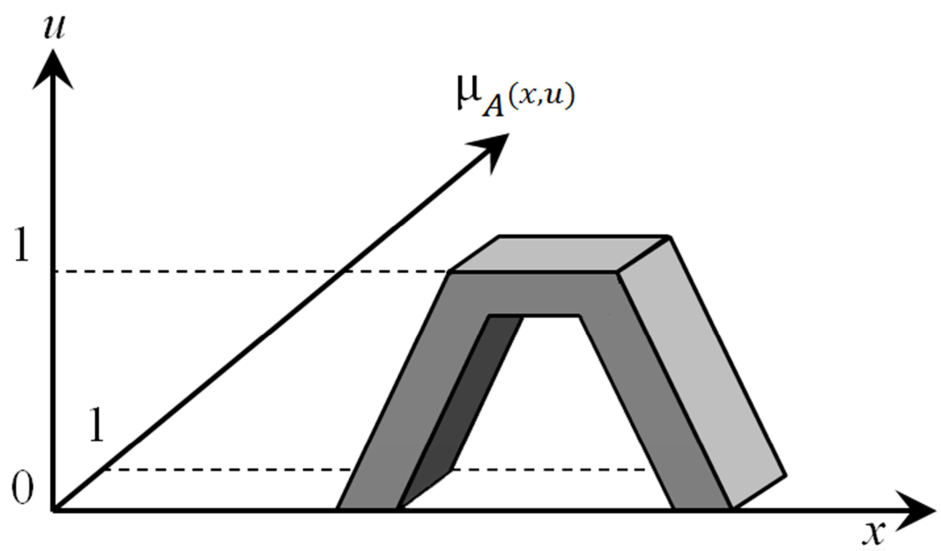


Figure 2.15: Example of an Interval Type-2 Fuzzy Set.

As shown in Figure 2.15, the difference between type-1 and type-2 fuzzy sets does not only lie in the third dimension since the two-dimensional part of the fuzzy set has also changed [Kassem 2012]. In order to explain this change, we provide an example of a trapezoidal type-1 fuzzy set A [Kassem 2012]. A trapezoidal fuzzy set is controlled by *4* parameters *a*, *b*, *c* and *d* as shown in Figure 2.16a. Now, if these parameters are blurred (fuzzy), i.e., exact values for these parameters cannot be given, then this will mean that these values lie somewhere between the parameters of A and

A’ as shown in Figure 2.16b [Kassem 2012]. What happens in case A and A’ are embedded into one set is shown in Figure 2.16c. Similarly to type-1 fuzzy sets, type-2 Fuzzy sets can take different shapes. Some of these shapes are shown in Figure 2.17 [Kassem 2012].

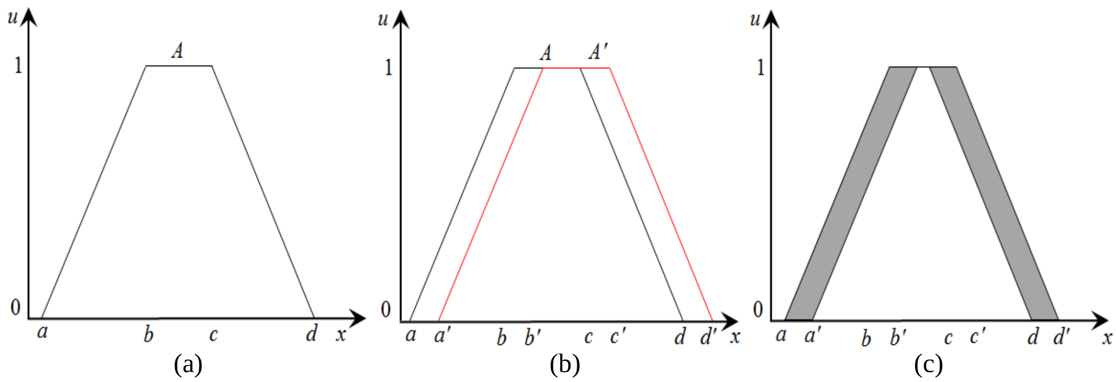


Figure 2.16: (a) Type-1 Trapezoidal Fuzzy Set, (b) Two Type-1 Trapezoidal Fuzzy Sets, (c) A Type-2 Trapezoidal Fuzzy Set.

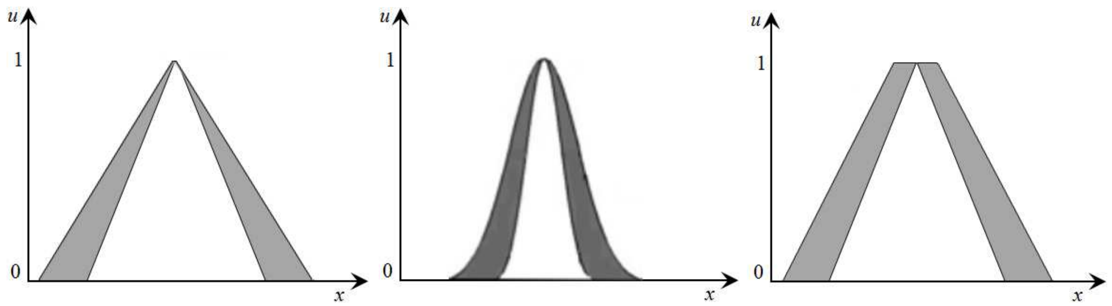


Figure 2.17: Some possible Type-2 Fuzzy Sets shapes.

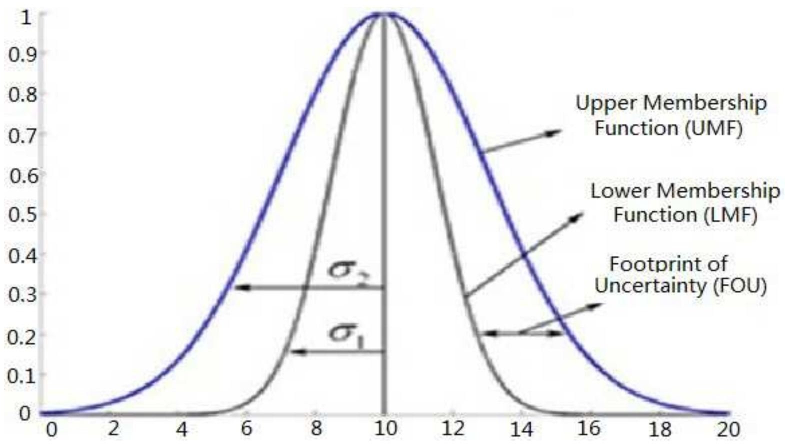


Figure 2.18 An interval type-2 fuzzy set.

An interval type-2 fuzzy set (IT2FS) A is characterized as [Hagras 2004], [Mendel 2006]:

$$A = \int_{x\in X\; u\in J_x \subseteq[0,1]} \int 1/(x,u) \;=\; \int_{x\in X}\left[\int_{u\in J_x\subseteq[0,1]} 1/u\right]/x \tag{2.23}$$

where x, the *primary variable*, has domain X; $u\in U$, the *secondary variable*, has domain J_X at each $x \in X$; J_X is called the *primary membership* of x and is defined in (5), and the *secondary grades* of A all equal 1. Note that (1) means: A : $X \rightarrow \{[a, b] : 0 \leq a \leq b \leq 1\}$. Uncertainty about A is conveyed by the union of all the primary memberships, which is called the *footprint of uncertainty* (FOU) of A, i.e.

$$FOU(A) = \cup_{\forall x\in X} J_x = \{(x,u): u \in J_x \subseteq [0,1]\} \tag{2.24}$$

The *upper membership function* (UMF) and *lower membership function* (LMF) of A are two type-1 MFs that bound the FOU (as shown in Figure 2.18). The UMF is associated with the upper bound of $FOU(A)$ and is denoted $\overline{\mu_A}(x)$, $\forall x\in X$, and the LMF is associated with the lower bound of $FOU(A)$ and is denoted $\underline{\mu_A}(x)$, $\forall x\in X$, i.e.

$$\overline{\mu_A}(x) \equiv \overline{FOU(\,A\,)} \qquad \forall x \in X \tag{2.25}$$

$$\underline{\mu_A}(x) \equiv \underline{FOU\ (\,A\,)} \qquad \forall x \in X \tag{2.26}$$

Note that J_X is an *interval set*, i.e.

$$J_x = \left\{(x,u) \colon u \in \left[\underline{\mu_A}(x)\,, \overline{\mu_A}(x)\right]\right\} \tag{2.27}$$

Theorem 1 (T2FS Representation Theorem Specialized to an IT2 FS): For an IT2 FS, for which *X* and *U* are discrete, the domain of *A* is equal to the union of all of its embedded T1 FSs, so that *A* can be expressed as

$$A = 1/FOU(A) = 1/\bigcup_{j=1}^{n_A} A_e^j = \sum_{i=1}^{N} u_i^j\,/x_i \tag{2.28}$$

The set theory operations of union, intersection and complement, which are widely used in applications of fuzzy sets, are especially easy to compute for IT2FSs. Given the IT2 FSs:

$$A = 1/FOU(A) = 1/\bigcup_{\forall x \in X}\left[\underline{\mu_A}(x)\,, \overline{\mu_A}(x)\right]\ ,\ B = 1/FOU(B) = 1/\bigcup_{\forall x \in X}\left[\underline{\mu_B}(x)\,, \overline{\mu_B}(x)\right] \tag{2.29}$$

then,

$$A \cup B = 1/\bigcup_{\forall x \in X}\left[\underline{\mu_A}(x) \vee \underline{\mu_B}(x)\,, \overline{\mu_A}(x) \vee \overline{\mu_B}(x)\right] \tag{2.30}$$

$$A \cap B = 1/\bigcup_{\forall x \in X}\left[\underline{\mu_A}(x) \wedge \underline{\mu_B}(x)\,, \overline{\mu_A}(x) \wedge \overline{\mu_B}(x)\right] \tag{2.31}$$

$$\overline{A} = 1/\bigcup_{\forall x \in X}\left[1 - \underline{\mu_A}(x)\,, 1 - \overline{\mu_A}(x)\right] \tag{2.32}$$

Note that for each value of *x* the intersection and union operations are referred to as the *meet* and *join* operations, respectively.

2.1.7.2 Procedures of Interval Type-2 Fuzzy Logic Systems

The IT2FLS depicted in Figure 2.19 [Hagras 2004] uses interval type-2 fuzzy sets [Mendel 2001] 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 [Hagras 2004], [Mendel 2001]. The use of interval type-2 FLS helps to simplify the computation.

The interval type-2 FLS works as follows [Hagras 2004], [Mendel 2001]. 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 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 combines the output sets and performs a centroid calculation which leads to type-1 fuzzy sets. The latter are called the type-reduced sets. There are different types of type-reduction methods. In this chapter, we will be using 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 [Hagras 2004], [Mendel 2001]. After the type-reduction process, the type-reduced sets are defuzzified to obtain crisp outputs.

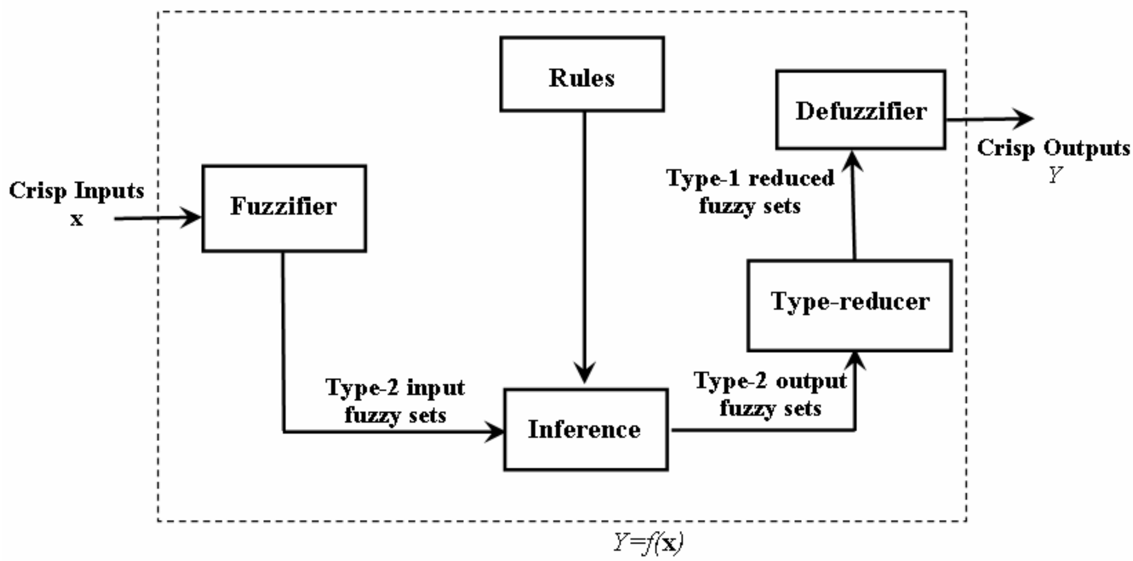


Figure 2.19 Structure of the type-2 FLS.

2.1.7.3 Singleton Fuzzification

The fuzzifier maps a crisp input vector with p inputs $x = (x_1 \dots x_p)^T \in X_1 \times X_2 \dots X_p \equiv X$ into input fuzzy sets; these fuzzy sets can on the whole be type-2 fuzzy input sets $\tilde{A}_x$ [Mendel 2001], [Hagras 2004]. However, we will use singleton fuzzification since it is fast to compute. In the singleton fuzzification, the input fuzzy set has only a single point of nonzero membership, i.e., $\tilde{A}_x$ is a type-2 fuzzy singleton if $\mu_{\tilde{A}_x}(x) = 1/1$ for $x = x'$ and $\mu_{\tilde{A}_x}(x) = 0$ for all other $x \neq x'$ [Mendel 2001].

2.1.7.4 Rule Base

According to [Mendel 2001], the structure of rules in a T1FLS and a T2FLS is the same. The difference is that in T2FLS the antecedents and the consequents will be represented by interval type-2 fuzzy sets. Therefore in a T2FLS for behaviour recognition there are p inputs $x_1 \in X_1, \dots, x_p \in X_p$ representing the p input behaviour features and c outputs $y_1 \in Y_1, \dots, y_p \in Y_c$ representing the occurring possibilities of the target behaviours. Suppose there are M rules then the i^{th} rule in this multiple-input-multiple-output (MIMO) FLS can be written as follows:

$$R^i_{MIMO}\text{: IF } x_1 \text{ is } \tilde{F}^i_1 \text{ …and } x_p \text{ is } \tilde{F}^i_p\text{, THEN } y_1 \text{ is } \tilde{G}^i_1 \text{ …and } y_c \text{ is } \tilde{G}^i_c,\ \ i = 1, \dots, M \tag{2.30}$$

According to [Mendel 2001], the R^i_{MIMO} can be considered as a group of multiple-input–single-output (MISO) rules R^i_{MISO}.

2.1.7.5 Inference

In the T2FLS the inference engine combines rules and provides a mapping from input type-2 fuzzy sets to output type-2 fuzzy sets [Hagras 2004], [Mendel 2001]. In the inference engine, multiple antecedents in the rules are connected using the *Meet* operation, the membership degrees in the input sets are combined with those in the output sets using the extended sup–star composition, and multiple rules are combined using the *Join* operation [Hagras 2004], [Mendel 2001]. In a MISO rule base, the *i*th rule can be written as follows [Hagras 2004], [Mendel 2001]:

$$R^i_{MISO}\colon \tilde{F}^i_1 \times \dots \times \tilde{F}^i_p \rightarrow \tilde{G}^i_k = \tilde{A}^i \rightarrow \tilde{G}^i_k,\ \ i = 1, \dots, M \tag{2.31}$$

R^i_{MISO} is represented by the membership function $\mu_{R^i}(x, y) = \mu_{R^i}\big(x_1, \dots, x_p, y\big)$,where

$$\mu_{R^i}(x, y) = \mu_{\tilde{A}^i \rightarrow \tilde{G}^i_k}(x, y) \tag{2.32}$$

can be written as [Hagras 2004], [Mendel 2001]:

$$\mu_{R^i}(x, y) = \mu_{\tilde{A}^i \rightarrow \tilde{G}^i_k}(x, y) = \mu_{\tilde{F}^i_1}(x_1) \sqcap \dots \sqcap \mu_{\tilde{F}^i_p}\big(x_p\big) \sqcap \mu_{\tilde{G}^i_k}(y) = \Big[\sqcap_{a=1}^{p} \mu_{\tilde{F}^i_a}(x_a)\Big]\, \mu_{\tilde{G}^i_k}(y) \tag{2.33}$$

In the T2FLS we will use *meet* under product t-norm so the result of the input and antecedent operations, which are contained in the firing set $\sqcap_{a=1}^{p} \mu_{\tilde{F}^i_a}(x'_a) \equiv F^i(x')$, is an interval type-1 set, as follows [Hagras 2004], [Mendel 2001]:

$$F^i(x') = \left[\underline{f}^i(x'), \overline{f}^i(x')\right] \equiv \left[\underline{f}^i, \overline{f}^i\right] \tag{2.34}$$

where $\underline{f}^i(x')$ and $\overline{f}^i(x')$ can be written as follows,

$$\underline{f}^i(x') = \underline{\mu}_{\tilde{F}_1^i}(x'_1) * \dots * \underline{\mu}_{\tilde{F}_p^i}(x'_p) \tag{2.35}$$

and

$$\overline{f}^i(x') = \overline{\mu}_{\tilde{F}_1^i}(x'_1) * \dots * \overline{\mu}_{\tilde{F}_1^i}(x'_p) \tag{2.36}$$

where * denotes the product operation.

2.1.7.6 Type Reduction

The type reduction returns type-1 fuzzy set outputs, which are then defuzzified to generate crisp outputs to represent the occurring possibility of the target behaviours in the current frame so that behaviour recognition can be performed. In our T2FLS for behaviour recognition, we used a centre of sets (cos) type-reduction, which has a reasonable computational complexity. The type-reduced set using the centre of sets type-reduction can be described in the following equation, and $Y_{cos}(x)_k$ for the k^{th} output which is an interval set determined by its left most point y_{lk} and its right most point y_{rk} can be expressed as [Hagras 2004], [Mendel 2001]

$$Y_{cos} = [y_{lk}, y_{rk}] = \int_{y_k^1 \in [y_{lk}^1, y_{rk}^1]} \cdots \int_{y_k^M \in [y_{lk}^M, y_{rk}^M]} \int_{f^1 \in [\underline{f}^1, \overline{f}^1]} \cdots \int_{f^M \in [\underline{f}^M, \overline{f}^M]} 1 \Big/ \frac{\sum_{i=1}^M f^i y_k^i}{\sum_{i=1}^M f^i} \tag{2.37}$$

To compute Y_{cos}, the centroids of the type-2 output sets y_k^t, which correspond to the rule consequents, must be calculated iteratively using the following equation in advance and before initializing the vision T2FLS recognition, where $z = 1, \dots, Z$, and Z is the iteration count, $t = 1, \dots, T$, and T is the amount of output fuzzy sets representing this output [Hagras 2004], [Mendel 2001].

$$y_k^t = [y_{lk}^t, y_{rk}^t] = \int_{\theta_1 \in J_{y1}} \dots \int_{\theta_z \in J_{yz}} 1 / \frac{\sum_{z=1}^{Z} y_z \theta_z}{\sum_{z=1}^{Z} \theta_z} \tag{2.38}$$

After that, we can perform the type-reduction by computing Y_{cos}. For this we need to compute the two end points y_{lk} and y_{rk} and then attach the firing strength f^i to the centroid of the *i*th rule consequent calculated in the previous step, as shown in the following equations [Hagras 2004], [Mendel 2001]:

$$y_{lk} = \frac{\sum_{i=1}^{M} f_l^i y_{lk}^i}{\sum_{i=1}^{M} f_l^i} \tag{2.39}$$

$$y_{rk} = \frac{\sum_{i=1}^{M} f_r^i y_{rk}^i}{\sum_{i=1}^{M} f_r^i} \tag{2.40}$$

2.1.7.7 Defuzzifier

Based on the Y_{cos} returned from the type-reduction, the defuzzification can be performed by using the average of y_{lk} and y_{rk} to get the crisp output for each output *k* [Hagras 2004], [Mendel 2001]:

$$Y_k(x) = \frac{y_{lk} + y_{rk}}{2} \tag{2.41}$$

2.2 Overview on Optimization Methods for Fuzzy Systems

In recent decades fuzzy logic systems have been successfully utilized to solve complex problems in various applications such as in classification (recognition) [Chi 1996], modelling [Pedrycz 1996], as well as control [Hagras 2004] in a significant number of applications [Hagras 2004] [Mendel 2001]. In most of these application cases, the importance factor for achieving successful accuracy and performance was the capability to incorporate human expert knowledge into fuzzy systems. However, complex applications usually come with large amount of parameters to model the problem so that the uncertainties can be handled. Therefore, it is not reasonable to only use human expert knowledge to decide and determine the parameters when building the fuzzy logic system. The main reason for this is that high-levels of uncertainties require much more adaptiveness so as to accommodate their changes in the real-world environments. In order to improve the learning capability and adaptiveness of the fuzzy logic system, two of the most successful methods are genetic algorithm (GA) and Big Bang-Big Crunch (BB-BC) optimization algorithm.

Overall, a genetic fuzzy system [Goldberg 1989] is a fuzzy logic system which was improved because of the enabling of the automatic learning capability by tuning the parameters of the fuzzy system using a genetic algorithm [Man 2000], which is a robust but slow learning method by imitating the natural evolution; in this way the system can have robust search capabilities in complicated problem spaces and thus provide a valid solution of the parameters to be tuned [Goldberg 1989], [Goldberg 2002], [Holland 1975]. The learning procedures of genetic algorithm involve different levels of complexity in relation to the evolving changes generated by the GA procedures and complex problem spaces [DeJong 1988]. Thus, parameter tuning and optimising has been one of the main approaches utilized to adapt a wide range of different fuzzy systems to the large amount of parameters associated with the complex problems by employing genetic fuzzy algorithms [Alcalá 2007a] [Alcalá 2007b] [Cordon 2004], [Casillas 2005]. The main disadvantage of these methods, which are based on genetic algorithm, is the extremely poor performance due to their slow convergence speed. This problem is especially obvious when solving complex problems which usually involve large number of parameters.

Another important optimization algorithm is the Big Bang-Big Crunch optimization. The BB-BC optimization is a heuristic population based on an evolutionary approach, which was presented by Erol and Eksin [Erol 2006]. 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. In [Erol 2006], comparisons between BB-BC against GA were performed. According to their comparison results, the performance of BB-BC exhibits superiority over an improved and enhanced genetic search algorithm, and BB-BC outperforms the GA for many benchmark test functions and comparison experiments with a much faster convergence speed. In [Kaveh 2010], BB-BC demonstrates better performance and outperforms the other optimisation algorithms such as genetic algorithms, evolution strategies algorithm, simulated annealing, tabu search, ant colony optimization, and harmony search. Similar comparison can be found in [Afshar 2011] which shows that BB-BC outperforms GA in their experiments. The reason for this fact is that, according to [Erol 2006], GA suffers from premature convergence, convergence speed and execution time problems in global optimum searching as they are generally sluggish in reaching the global optimum accurately and reliably in a short period of time; and by contrast, BB-BC finds the exact global optimum point for the sphere, step, Rastrigin functions within the maximum number of allowed iterations and outperforms the GA method The BB-BC theory is formed by two phrases: a Big Bang phrase where candidate solutions are randomly distributed over the search space in a uniform manner [Kumbasar 2011] and a Big Crunch phrase where candidate solutions are drawn into a single representative point via a centre of mass or minimal cost approach [Erol 2006]. All subsequent Big Bang phases are randomly distributed to the centre of mass or the best fit individual in a similar fashion. The procedures of the BB-BC are as follows [Kumbasar 2011]:

- Step 1 (Big Bang Phase): An initial generation of N candidates is generated randomly 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 centre of mass is chosen as the centre point. The centre of mass is calculated as:

$$x_c = \frac{\sum_{i=1}^{N} \frac{x_i}{f^i}}{\sum_{i=1}^{N} \frac{1}{f^i}} \tag{2.42}$$

where x_c is the position of the centre of mass, x_i is the position of the candidate, f^i is the cost function value of the ith 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:

$$x^{new} = x_c + \frac{\gamma\rho(x_{max} - x_{min})}{k} \tag{2.43}$$

where γ is a random number, ρ is a parameter limiting search space, x_{min} and x_{max} are the lower and upper limits, and k is the iteration step.

- Step 5: Return to Step 2 until stopping criteria have been met. Examples stopping criteria are: (1) current iteration number equals to the maximum iteration number; (2) the error (the difference between the current actual output and the expected output) is lower than a threshold value; (3) the accumulated running (consumed) time of the entire optimization procedure is larger than the given time for example five hours.

2.3 Overview on Human Event and Behaviour Detection

Detecting human events and activities from a video is one of the most promising applications of video processing. This research area has recently caught the attention of researchers from industry, academia, security agencies, and consumer agencies. The problem of human behaviour recognition has been a widely studied subject in the computer vision literature [Weinland 2010], [Aggarwal 2011]. In most of the existing methods, the first procedure is the feature extraction, which is used to describe the characteristics of the subject, and where the methods can be roughly classified into four categories: motion-based [Efros 2003], [Fathi 2008], [Wang 2008], [Wang 2007], appearance-based [Elgammal 2003], [Thurau 2008], space-time volume-based [Blank 2005], [Ke 2007], [Laptev 2007], space-time interest points and local feature-based [Dollar 2005], [Laptev 2008], [Niebles 2007], [Nowozin 2007], [Schuldt 2004]. Behaviour recognition approaches are mostly based on machine learning techniques employed in the pattern recognition literature. This includes techniques such as k-Nearest Neighbour (k-NN) [Efros 2003], [Wang 2007], [Niebles 2007], [Blank 2005], [Thurau 2008], [Weinland 2008], Support Vector Machine (SVM) [Dollar 2005], [Laptev 2008], [Schuldt 2004], [Jhuang 2007], [Liu 2008], [Schindler 2008], boosting-based classifiers [Nowozin 2007], [Fathi 2008], [Laptev 2007], Hidden Markov Model (HMM) [Elgammal 2003], [Vezzani 2010], [Yamato 1992]. In [Blank 2005], M. Blank et al. proposed a feature set of pose primitives for behaviour representation and n-Gram models which were utilised for behaviour matching and recognition. In [Wang 2007], Y. Wang et al. developed a feature set modelling behaviour as a set of the lowest distances from exemplars to behaviour images in an exemplar-based space. Yamato et al. [Yamato 1992] used discrete HMMs to recognise image sequences of six different tennis strokes among three subjects. Vezzani et al. in [Vezzani 2010] proposed an HMM-based action recognition method using model based feature set. However, it remains a challenging problem to achieve an automatic behaviour classification due to the huge complexity and uncertainties of the dynamic environments such as complicated scene background, occlusion, and varying postures and sizes of moving objects. To handle the uncertainties from the real-world environments, Type-1 Fuzzy Logic Systems (T1FLSs) have been applied to human activities analysis [Anderson 2008], [Anderson 2009] which performs well in predefined situations such as a quick change of the silhouette orientation. However, they lack general applicability and they always require time-consuming multi-camera calibration and manual operations. In order to improve the robustness of T1FLSs in handling uncertainties in the fields of image processing and pattern recognition, type-2 fuzzy logic systems have been proposed and reported to have successfully enhanced the reliability of the type-1 fuzzy logic systems [Almohammadi 2014], [Almohammadi 2015], [Yao 2012], [Yao 2014].

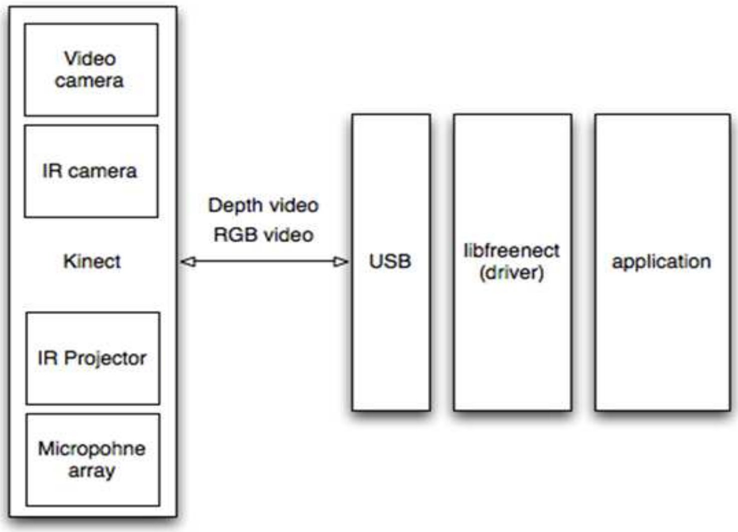
In this thesos, we will present a robust framework for machine vision based on human behaviour recognition using Interval Type-2 Fuzzy Logic Systems (IT2FLSs). In order to obtain the optimized parameters of the membership functions and rule base of the IT2FLS, we employed an optimization approach based on the Big Bang–Big Crunch (BB-BC) [Erol 2006], [Kumbasar 2011] algorithm. We will present several experiments which were performed on the publicly available Weizmann human action dataset. It will be shown that the proposed IT2FLS outperformed its T1FLS counterpart as well as other traditional non-fuzzy systems.

2.4 Overview on 3D Human Daily Activity Recognition

Robustly recognizing people's daily activities (e.g., sitting on a sofa, reading a book, falling down, drinking, and eating food, etc.) in a low-cost and automatic way (e.g., vision-based) for elders and children who are alone in a house is essential for providing them with proper healthcare and medical services. Vision-based human activity analysis has been one of the most active research topics in machine vision over the past decades. However, the limitation of the sensors (i.e., 2D colour camera) is that it restricts traditional methods [Bobick 2001] in that they are barely capable to detect simple behaviour in a not very robust way. However, since human bodies and behaviour are in essence of a three-dimensional structure, the loss of information and feature cues in the depth feature leads to a huge degradation of the demonstration for those 2D feature models [Efros 2003], [Black 1997], [Ni 2009].

To solve this problem, the recent presentation of depth sensor (e.g., Microsoft Kinect v1, as shown in Figure 2.19, has been presented and has become a feasible and reliable 3D sensor which is able to capture both the colour images (by using the video RGB camera) and depth information (by its 3D depth sensor) with appropriate resolution and accuracy in real time speed. It can offer the users the three-dimensional structure of the monitoring background as well as the three-dimensional motion movement of the human/objects in the view. Therefore the human behaviour analysis and daily activity recognition can be performed reliably and robustly by using this 3D visual sensor [Ni 2013].

(a)



(b)

Figure 2.20: The Structure of 3D Sensor. (a)The Structure of Kinect v1, (b) Components Diagram of Kinect v1.

In 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 layman's terms the information of interest within a long video sequences recorded at such spaces. Such summarization can be used to automatically detect serious events that need immediate attention such as attempted burglaries, serious injuries, etc. A 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 a day to avoid dehydration. Similarly, a warning message can be sent to social services nearby in case of an elderly person falling so that proper actions can be taken instantly. Likewise, lights can be turned off automatically when a user is detected to be sleeping on the sofa.

Various studies have been conducted for the linguistic summarization of video sequences where type-1 fuzzy logic systems have been applied for linguistic summarization and activities analysis. For example, in [Anderson 2008], [Anderson 2009] voxel person analysis was implemented 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, they require time-consuming multi-camera calibration (when a 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 [Trivino 2008] using wearable devices to summarize and analyse the human activity. However, such wearable devices are intrusive and can be inconvenient for the users.

As a key procedure in linguistic summarization, analysis of human behaviour and activity has attracted a great deal of scholarly interest. Most previous research on behaviour and activity recognition is based on 2D video data [Aggarwal 2011], [Ryoo 2009] or RFID sensors [Wu 2007]. However, the use of 2D data in real-life circumstances has 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 humans or on the objects. One traditional method is to employ spatio-temporal features to describe points of interest in 2D video data [Laptev 2005]. This method can be later improved by adding more information to model the features [Jhuang 2007]. There have been various works for activity analysis employing the Hidden Markov Model (HMM), which was firstly used to analyse two-hand behaviour [Brand 1997]. Later HMM was used to recognize gesture and posture through a probabilistic framework [Elgammal 2003]. However, the accuracy was not satisfactory. Dynamic Time Warping (DTW) is another method to measure similarity and distance between two behaviours [Zhou 2008]. However, DTW only returns exact values and thus it is inadequate to model the uncertainty and ambiguity of behaviour.

In this thesis, we will present a robust behaviour recognition algorithm for video linguistic summarization using a 3D Kinect camera based on Interval Type-2 Fuzzy Logic Systems. 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) algorithm. Our experiments have been successfully conducted in real-world intelligent environments and the 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 are presented including video linguistic summarizations event searching and activity retrieval/playback.

2.5 Discussion

At the beginning of this chapter, we introduced fuzzy sets and the concept of fuzzy logic. Overall, there are type-1 fuzzy logic systems and type-2 fuzzy logic systems. The major problem of type-1 fuzzy logic system is that it is incapable of robustly modelling the high levels of complex uncertainties associated with real-world environments. To solve this problem, interval type-2 fuzzy logic system was proposed and widely used. The main advance extension of IT2FLS is that it utilizes type-2 fuzzy sets which have proven to be very robust in handling the uncertainties when compared to type-1 fuzzy sets. In this chapter, we also elaborated on the interval type-2 fuzzy logic system and the modules that distinguish it from the type-1 fuzzy logic system.

Since an interval type-2 fuzzy logic system is more capable of robustly handling the complex uncertainty there are also more parameters involved in interval type-2 fuzzy logic system than in type-1 fuzzy logic system. To achieve a robust performance, these parameters need to be decided on and tuned in a reasonable paradigm so that the adaptiveness of the system is ensured.

We presented the basis of Big Bang-Big Crunch which a heuristic population based on an evolutionary approach that is derived from one of the theories of the evolution of the universe in physics and astronomy, namely the BB-BC theory. The main advantages of Big Bang-Big Crunch optimization algorithm are its low computational cost, ease of implementation, and fast convergence. Unlike classic conventional optimization algorithm genetic algorithm that needs days or weeks to finish the convergence, Big Bang-Big Crunch optimization algorithm normally achieves the convergence in hours or even minutes. This advantage is highly important for intelligent systems in solving the modern problems which are usually complicated by hundreds or even thousands of parameters. Big Bang-Big Crunch models possible solutions to this problem by using a population structure which is composed of parameters to be tuned. In the initialization stage of Big Bang-Big Crunch, we can produce a random set of populations and we can also use sophisticated algorithms such as Fuzzy C-Means and Wang-Mendel to calculate a reasonable starting point (which will be discussed in detail in the rest of the chapters in this thesis) of the optimization algorithm so that Big Bang-Big Crunch can converge to the ideal solution more quickly and accurately. After that the Big Bang-Big Crunch assesses each population (candidate solution) via a fitness function to compute a score value for each candidate solution describing how good the solution is and then save the best solution while regenerating the other population. The Big Bang-Big Crunch then goes through this procedure in an iterative way. Each iteration determines whether the current optimization has satisfied the stopping criterion, and therefore, this iterative optimization procedure drives the population towards better solutions by keeping only the best solution with the highest fitness score from one generation to another.

In this chapter, we briefly elaborated on human behaviour recognition and its common steps. Generally speaking, the first step is to perform a background/foreground segmentation so that we can define the areas and locations of the moving targets in the current frame. For most applications of the smart vision system, the main analysis targets are human users. However, due to the complexity of the real-world environments, the obtained foreground contexts are not only human users but also various non-human objects. Therefore, the next step involves the human silhouette extraction based on the foreground images. There are several methods which have been reported regarding this in recent years, and these conventional methods are able to extraction human silhouette. But there are still problems in these traditional algorithms such as an extracted human silhouette being degraded due to misclassifications caused by the noise factors and uncertainties. Our proposed method, which is based on interval type-2 fuzzy logic system, can solve this problem since the type-2 fuzzy logic can accommodate more possibilities in the solution space and model more uncertainty. As a result, performance and accuracy are improved in comparison to when the conventional methods are applied. Once we obtained the silhouette, we deciphered the area of the human subject in the current frame. Then the next step was to build a model to describe the behavioural features based on which we could classify the current action. After that we further developed the conventional algorithms for behaviour recognition and the type-1 fuzzy logic based methods with the aim of solving this problem. To improve the accuracy and to achieve better recognition performance in the high-levels of uncertainties, we introduced our proposed recognition method based on interval type-2 fuzzy logic system.

In the last part of this chapter, we introduced a system framework of behaviour recognition and event summarization using 3D sensor in real-world environments. Choosing the proper sensor according to the actual application scenario is always the first and also a very important step in designing the application system since different types of sensors have different advantages and disadvantages. In the scenario of ambient assisted living environments, the main problems are the high-level of uncertainties caused by the casualty of the human behaviour, the occlusion problem of the multiple crowded subjects and the various noise factors existing in the environments. There are two solutions for AAL. One is based on a 2D video sensor but the problem is that it is not able to achieve an accurate recognition since a 2D video sensor cannot capture robust features against so many uncertainties and noise factors. The other solution involves wearable sensors. This is a robust solution but it is also too intrusive for the users as the wearable sensors are usually attached to the skin or muscle of the human users. We therefore proposed a system framework for AAL based on a 3D sensor which is very robust and it is not wearable. We used an interval type-2 fuzzy logic system based recognition to handle the high-level of uncertainties during the recognition procedure.

In the following chapter, we will mainly discuss the algorithms of human silhouette extraction.

Chapter 3: A IT2FLS Based Approach for Human Silhouette Extraction in Intelligent Environments

3.1 Introduction

In this chapter, we will explain the fundamental but important module of vision-based AAL environments, human silhouette extraction, which detects the area and location of the subject in the view. This provides the basic feature information for further higher-level modules such as human tracking, behaviour recognition and event summarisation. There are several traditional methods for silhouette extraction. However, several noise factors (shadow, reflection, obstacles attached to the human subjects, etc.) exist in real world environments. To address this problem, a T1FLS-based algorithm was proposed and was able to effectively eliminate the noise factors during the procedure of silhouette extraction. But due the high-levels of uncertainty, this method has a problem of silhouette degradation. In this chapter, in order to solve this problem, we will propose an IT2FLS-based method for silhouette extraction which is capable of robustly extracting the silhouette and eliminating the noise factors.

3.2 Background of Human Silhouette Extraction

Automatic human event detection is an essential prerequisite for many advanced applications, such as smart home, visual surveillance, pervasive computing. In this research, we are interested in automatically detecting human event from video stream for smart inhabited environments [Callaghan 2004]. For human event detection, the first procedure is to segment (or outline) a human silhouette from a video sequence. Since silhouettes in images and videos give important information about the foreground objects that can be utilized for a high level analysis, silhouettes are usually critical hints of an image and video understanding and will impact the robustness of human event detection significantly. However, in real-life environments, for example in a smart living-room monitoring, due to the huge complexity of such environment, there are numerous noise factors and uncertainties present for silhouette extraction and event detection which include:

- Varying light condition
- Reflections and shadows
- Moving objects attached to human silhouette (a book, a chair, etc.).

Advanced human detection and identification approaches like [Dalal 2005], [Wu 2006] can be utilized for silhouette extraction. However, such methods are commonly of high computational complexity and hence not suitable for dynamic and complex environments. Hence, there is a need for efficient silhouette extraction methods and that are able to operate in dynamic environments. Therefore, in this research, and as a fundamental part of our project whose long-term vision is to construct a large scale intelligent system that can automatically detect human behaviour and event to provide related services, we will propose an effective and efficient way to extract human silhouette and to further detect particular human events. Specifically, CCTV cameras have been deployed in our laboratory of smart living room to capture video data of the users' regular activity. Via these cameras, we analyse the occupants' behaviour and detect particular events.

The first step of detecting a human event is a background subtraction for silhouette extraction to detect moving targets as foreground objects. In [Aubert 2001], an approach based on a single Gaussian modal was developed which employed a simple adaptive method. However, there are several limitations in this method such as learning stage necessity for background distribution, robustness deficiency for situations like sudden illumination changes, slow moving objects, and so on. To address these problems, the Gaussian Mixture Model (GMM) was proposed [Stauffer 1999] [Katz 2003]. Since GMM is extensively recognized as a robust method for background subtraction, in this chapter, GMM is utilized for foreground detection. However, it is unreasonable to simply consider GMM foreground as human silhouette in real-life environments due to the noise factors mentioned above. For example, GMM may segment objects (a book, a chair, etc.) moved or held by a person as part of the foreground which will cause noise to further procedures such as event detection module and thus lead to false alarms.

To handle these problems and detach the moving objects from the human silhouette, a type-1 Fuzzy Logic System (T1FLS) was proposed [Chen 2006]. T1FLS is capable of handling to an extent the uncertainties mentioned above, however, the extracted silhouette will be degraded due to misclassification of the proposed T1FLS. Hence, here we present an Interval Type-2 Fuzzy Logic System (IT2FLS) which will be able to handle the high uncertainty levels present in real-world dynamic environments while also reducing the misclassification of extracted silhouette. The IT2FLS uses a similar type-1 membership function as the ones presented in [Mendel 2001] as principal membership functions which are then blurred to produce the type-2 fuzzy sets used in this chapter. We also used the same rule base as [Chen 2006] to allow for a fair comparison with the results reported in [Chen 2006].

In this proposed system, GMM is adopted to detect original foreground which may include a moving object such as a cup, a book moved or held by human. Then, in order to eliminate these noise factors and extract a pure human silhouette, an IT2FLS is performed to detach the moving objects from the human silhouette. Several

Variables such as SAD, distance, neighbours are calculated for the IT2FLS to get the silhouette degree of each pixel (or image block) to determine whether it belongs to the human silhouette. Finally, pixels (or image blocks) having a higher silhouette degree than 50% are considered as human silhouette. We performed several real-world experiments where it was shown that the proposed IT2FLS is effective in reducing the misclassification and that the quality of the extracted human silhouette is much improved when compared to the T1FLS.

3.3 The Proposed Fuzzy based Human Silhouette Extraction

Figure 3.1 provides an overview of the proposed methodology for human based event detection in real-world public spaces. In the first stage, source images are captured using a video camera. The images are then analysed using GMM to detect the foreground. The foreground detected by GMM is then refined to eliminate noise factors and detach moving objects by utilizing several feature variables and human centroids obtained by human tracking module which is performed by global nearest neighbour (GNN) [Blackman 1999]. In this way, refined human silhouette is achieved which is then analysed to recognize the semantics of human behaviour. Finally, event inference is carried out by conjointly considering the behavioural semantics and human trajectories to detect particular human events.

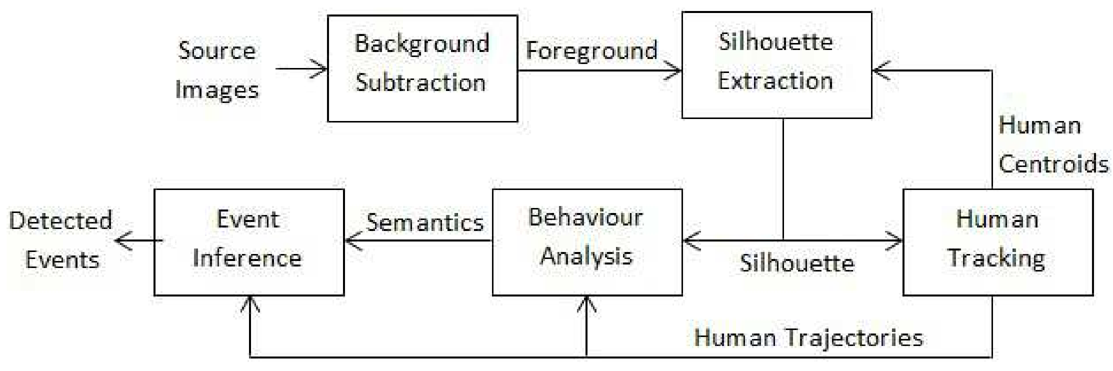


Figure 3.21: An overview of the proposed human based event detection methodology.

Among these stages, silhouette extraction is an important and fundamental procedure which greatly impacts the accuracy of the subsequent stages. However due to the complexity of the environment, numerous noise factors and uncertainties exist during this procedure. Indeed, advanced human detection and identification approaches can be utilized for silhouette extraction, but those methods are commonly of high computational complexity and not adaptive for dynamic environments. Hence in [Chen 2006] a T1FLS was presented while in this chapter an IT2FLS for silhouette extraction will be presented.

Suppose that we are working on frame *i*, and the foreground image of frame *i* obtained by GMM is denoted by O_i, and the silhouettes in frame *i*-1 that have been properly segmented are denoted by O_{i-1}. As described above, the foreground in O_i, may contain the human body and moving non-human objects attached to silhouette. To detach the moving objects and refine the human silhouette, the IT2FLS is developed based on the following observations:

1. If an image block in O_i belongs to a human body, it is of high probability to match an image block in O_{i-1} in a good match degree. The SAD (sum of absolute difference) between their corresponding blocks in frame *i* and frame *i*-1 is used to measure the matching degree between the image block in O_i and its best match block in O_{i-1}.
2. If the distance between the left-top pixel of this block and human centroid is large, the probability that this block belongs to the human body is low.
3. If the amount of the neighbouring block with high probability belonging to human body is huge, such as having good matches in O_{i-1}, or having low

distance to human centroids etc., then, the probability of this block also belonging to the human body is high.

Based on the observations above, the following variables of each block are calculated.

- SAD of motion estimation. For every image block in O_i, its best match block in frame *i*-1 is searched. The matching degree is used to describe the SAD variable.
- The distance between left-top pixel of this block and the human centroid.
- The amount of its neighbourhood with a high probability of belonging to human body, for example, by having a good match block in human body, low centroid distance, and so on.

The rules of the type-2 fuzzy system should remain the same as the T1FLS in [Chen 2006]. Moreover, in our IT2FLS we will use a similar type-1 fuzzy sets as those in [Chen 2006] as the principal membership functions which can be then blurred by α% (α = *10*, *20*, *30*, *40…*) to produce the Footprint of Uncertainty (FOU) of the corresponding type-2 fuzzy sets. The impact of different FOU to our proposed IT2FLS was analysed and the results are shown in Figure 3.2. The membership functions for the inputs and output of the IT2FLS are shown in Figure 3.2. The rule base of the IT2FLS is the same as [Chen 2006] and it is as follows:

1. If SAD is very low AND Neighbourhood is huge AND Distance is close,

   THEN Silhouette is high.

2. If SAD is large AND Neighbourhood is small AND Distance is very large,

   THEN Silhouette is low.

3. If SAD is low AND Neighbourhood is large AND Distance is medium,

   THEN Silhouette is high.

4. If SAD is medium AND Neighbourhood is medium AND Distance is medium,

   THEN Silhouette is medium.

5. If SAD is large AND Neighbourhood is medium AND Distance is medium,

   THEN Silhouette is medium.

6. If SAD is large AND Neighbourhood is large AND Distance is close,

   THEN Silhouette is high.

7. If SAD is medium AND Neighbourhood is large AND Distance is medium,

   THEN Silhouette is high.

8. If SAD is medium AND Neighbourhood is large AND Distance is close,

   THEN Silhouette is high.

9. If SAD is very large AND Neighbourhood is small AND Distance is large,

   THEN Silhouette is low.

10. If SAD is medium AND Neighbourhood is small AND Distance is very large,

    THEN Silhouette is low.

11. If SAD is low AND Neighbourhood is huge AND Distance is medium,

    THEN Silhouette is high.

12. If SAD is low AND Neighbourhood is medium AND Distance is medium,

    THEN Silhouette is high.

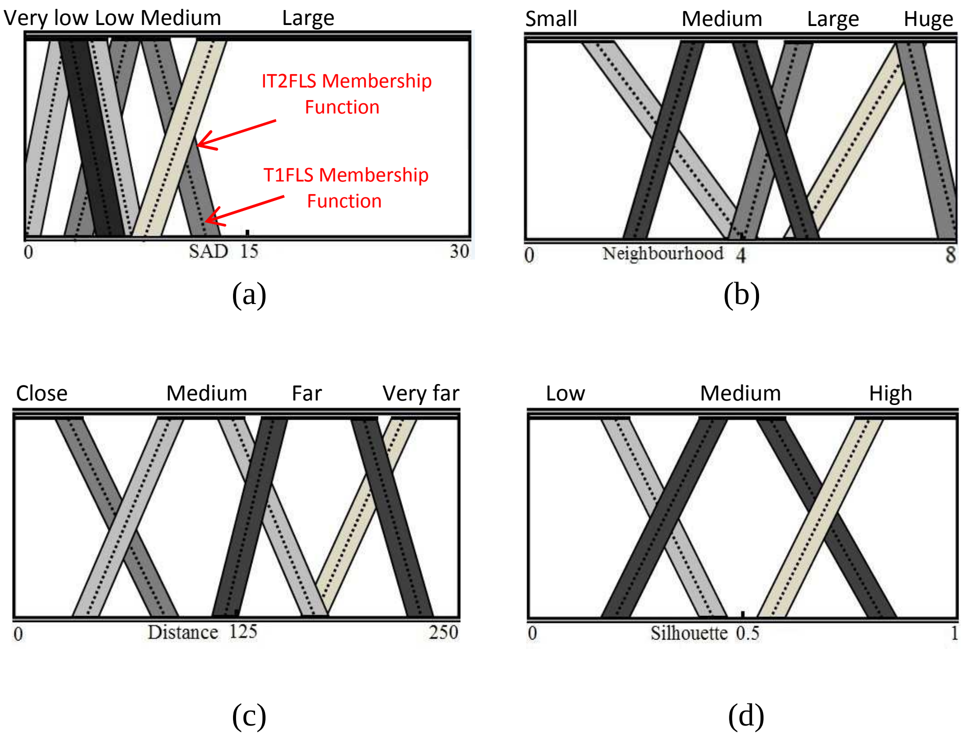


Figure 3.22: The interval type-2 fuzzy sets employed in the inputs of our IT2FLS for (a) SAD, (b) Neighbourhood, (c) Distance, (d) The Silhouette output of IT2FLS.

3.4 Experimental Results

We performed several real-world experiments to validate our proposed approach and to compare the performance of the proposed IT2FLS and the T1FLS presented in [Chen 2006]. The aim of the experiments was to confirm that the proposed IT2 fuzzy system is effective to detach the moving objects from human silhouette with much fewer misclassifications than a T1 fuzzy system.

In our experiments of this chapter, twelve human subjects were involved in four video sequences. And we performed four sessions run to analyse to the results. In the experiment of detaching the moving objects, instructed behaviour was used while free behaviour was in other tests of this chapter.

As shown in Figures 3.3 to 3.7, (a) shows the source images; (b) shows the original foreground detected by GMM; (c) illustrates results after the T1FLS and (d) exhibits results after using IT2FLS. In our case, a human silhouette is represented by pixels having a higher degree than the 0.5 degree of the fuzzy silhouette.

In Figure 3.3, the results of "Raising a book" demonstrate that the book attached to human is eliminated after using the fuzzy based systems. The experiment shows that two fuzzy systems extract a proper silhouette and that they are able to detach the book from human body. However, as mentioned above, the silhouette extracted by T1FLS is degraded due to a misclassification (as confirmed by [Chen 2006]) while IT2FLS can address this problem and reduce the misclassification.

In Figure 3.4 enlarged silhouette images of "Raising a book" are provided. As can be seen in Figure 3.4b, the edge the silhouette of T1FLS, (highlighted with red rectangles) is degraded to a coarser edge due to misclassification when compared to the original foreground. However, as displayed in Figure 3.4c, the IT2FLS with the uncertainty factor of $\alpha=20\%$ achieves the same human silhouette as the original foreground while the book is detached.

To demonstrate the robustness of the proposed system, more experiments were conducted in various environments. In Figure 3.5, the results obtained from an outdoor environment of single pedestrian under snowy conditions are shown. Through this experiment, we can see that the proposed system is working effectively in an outdoor environment. In the images of figure 3.6, one can see the reflection of human body where a noise factor is detected as foreground by GMM while the reflection is eliminated by utilizing fuzzy logic. In Figure 3.7, the results obtained from an outdoor environment crowded with people are shown. In this complex outdoor environment with more noise and uncertainties, the proposed system demonstrates its robustness by extracting the human silhouette with a promising result. In relation to these experiments and for the purpose of comparison, Table 3.1 provides the average misclassification and accuracy for T1 and IT2 fuzzy systems. In Table 3.1, it is clearly demonstrated that the misclassification of the proposed IT2FLS is significantly reduced compared to the T1FLS while the IT2FLS also results in a higher accuracy than the T1FLS.

One fact in choosing the block size for the SAD motion estimation is that the block size needs to be properly tuned and chose according to the video resolution and the size of the target subject. If the block size is set to a high value, for example, 64×64 in a VGA video (640×480), the accuracy of SAD motion estimation will degrade and further lower the robustness of the entire system, even though the speed of the entire system would be higher compared to a smaller block size usage.

The accuracy is obtained using the following paradigm. Suppose we are working on the refined silhouette images processed by fuzzy logic, in which noise factors are eliminated and moving objects are detached while degradation exists. We then compare the extracted human silhouette with their corresponding human targets foreground in the original foreground images and the number of misclassified pixels can be obtained because there is no degradation in the original foreground images. The misclassification is defined by the difference between the pixel count of original moving object O_i and its according silhouette object S_i (The number of missing pixels on the silhouettes). And the accuracy is defined by calculating the correct percentage of the pixel number of the silhouette objects comparing with its according original moving object O_i. After that, the accuracy can be calculated:

$$\text{Accuracy} = 1 - \sum_{i=1}^{n} (O_i - S_i) / \sum_{i=1}^{n} O_i \qquad (3.1)$$

$$\text{Misclassifications} = \sum_{i=1}^{n} (O_i - S_i) \qquad (3.2)$$

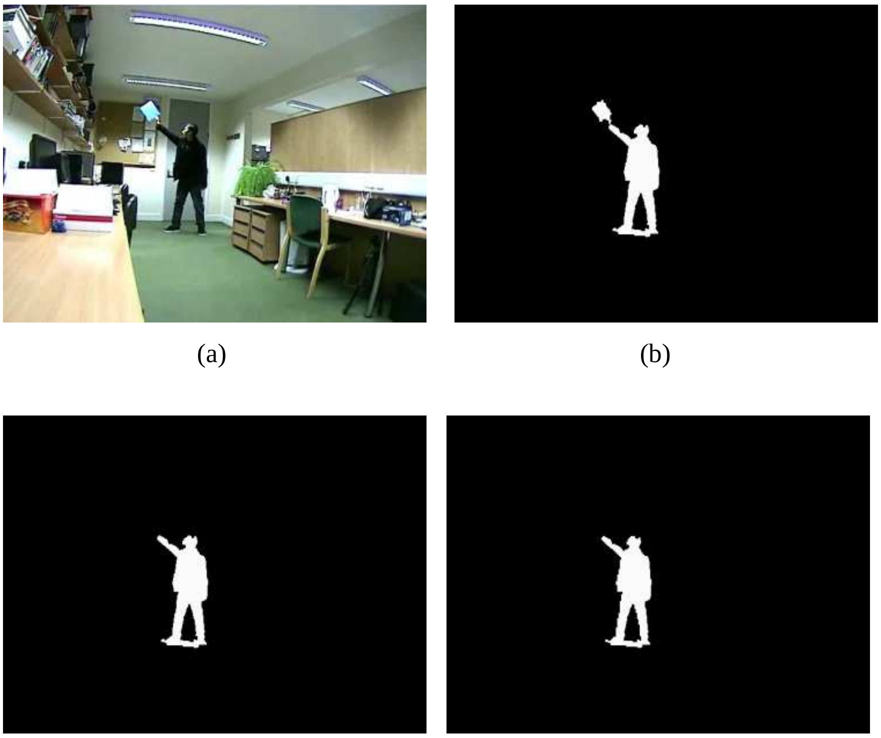
Where n is the number of human silhouette targets in the fuzzy refined silhouette images, and S_i represents the pixel count of extracted human silhouette target in fuzzy silhouette images while O_i denotes the pixel count of the corresponding human target of S_i in original foreground images.

Figure 3.8a shows that for the experiment involving multiple people in an open real-world public space (see Figure 3.8), the misclassification of T1FLS fluctuates a lot with more video frames used, while IT2FLS remains stable near to 0 (thus indicating nearly zero misclassification). In Figure 3.8b, it can be seen that the accuracy of the T1FLS against the number of video frames demonstrates fluctuation between 92.5% and 95.7% while the accuracy of our proposed IT2FLS stabilizes at around 99.99%. This shows the robustness of the employed IT2FLS.

From Figure 3.9 it is obvious that the best IT2FLS results in terms of the least average misclassification (Figure 3.9a) and highest accuracy are obtained for the uncertainty factor of $\alpha=20\%$. This means that the corresponding FOU to an uncertainty factor $\alpha=20$ % was the appropriate FOU to handle the encountered uncertainties in the given environment.

Experiment Name	T1 Avg Accuracy	T1 Avg Misclassify	T2 Avg Accuracy	T2 Avg Misclassify	Frame Used
1. Single person in outdoor snow environment	93.23%	125.74 pixels	99.9791%	0.61 pixels	383
2.1 Multiple person with reflection in outdoor clear environment	95.23%	221.38 pixels	99.9951%	0.54 pixels	246
2.2 Multiple person in outdoor snow environment	86.59%	393.86 pixels	99.8960%	6.4 pixels	386
3. Crowded outdoor environment in snow condition	90.21%	637.62 pixels	99.9004%	12.84 pixels	255

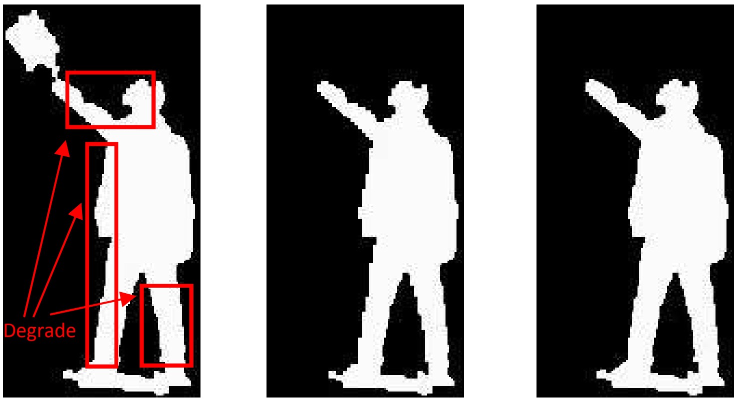
Table 3.1: Comparison of Misclassification and Accuracy



(a) (b)

(c) (d)

Figure 3.23: The experiment of “Raising a book in indoor environment”; (a) source images, (b) foreground detected by GMM, (c) extracted silhouette after using T1FLS, (d) extracted silhouette after using IT2FLS.



(a) (b) (c)

Figure 3.24: Enlarged images of “Raising a book in indoor environment”; (a) foreground detected by GMM, (b) extracted silhouette after using T1FLS, (c) extracted silhouette after using IT2FLS.

(a) (b)

(c) (d)

Figure 3.25: The experiment of “Single person in outdoor snowy environment”.

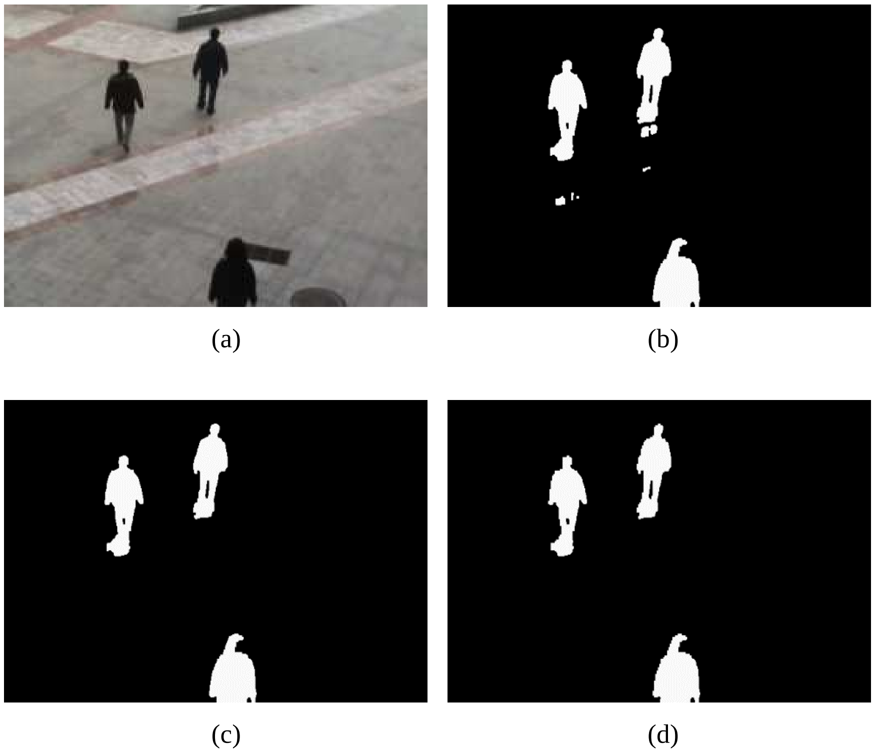


Figure 3.26: The experiment of “Multi-person with reflection in outdoor environment”.

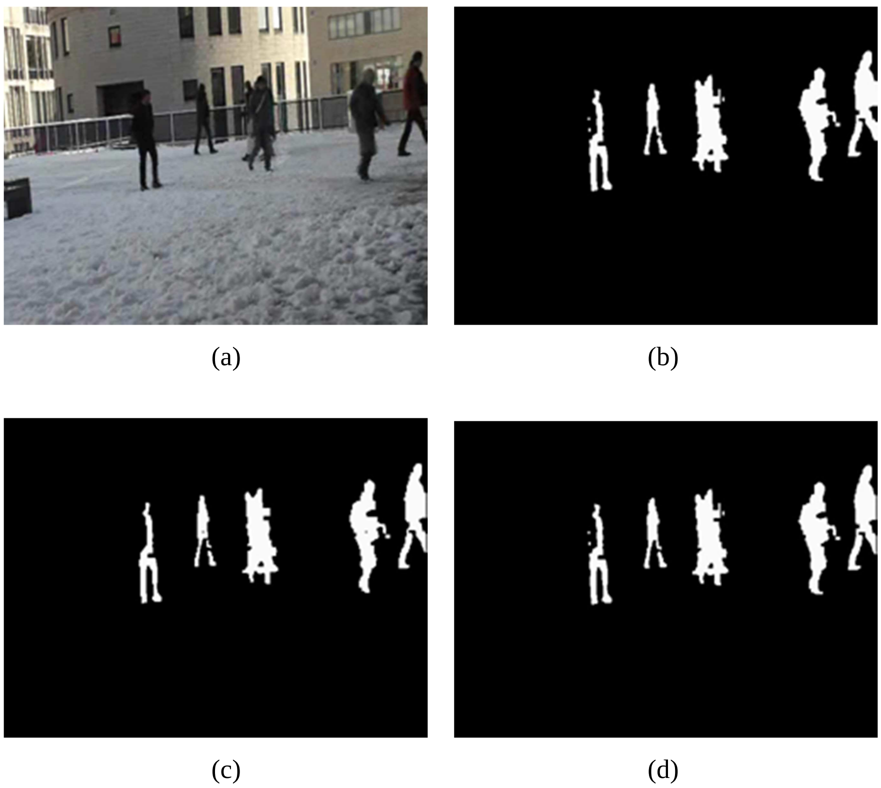
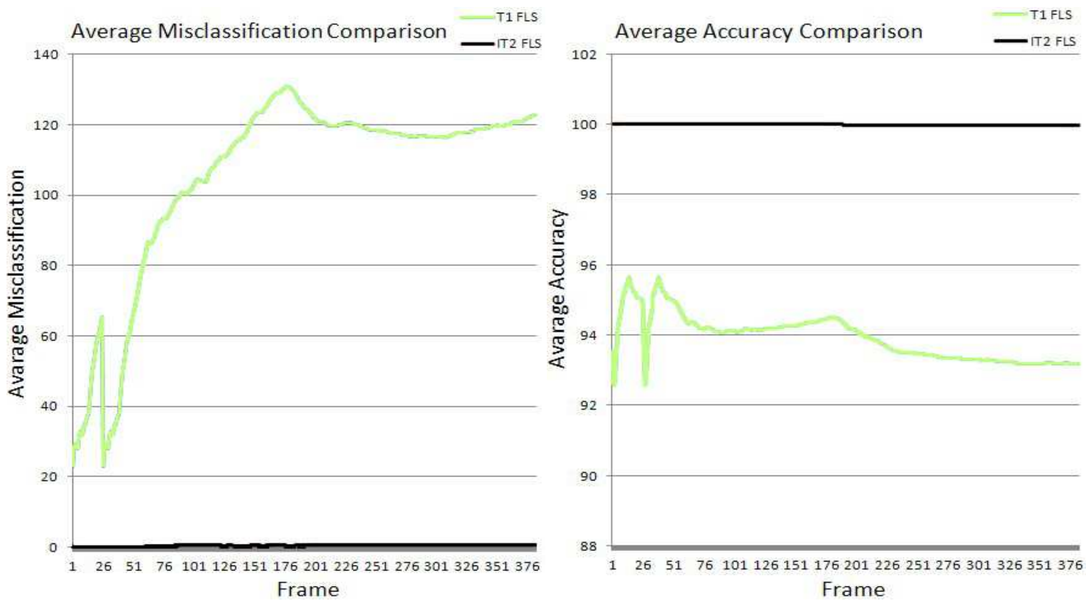
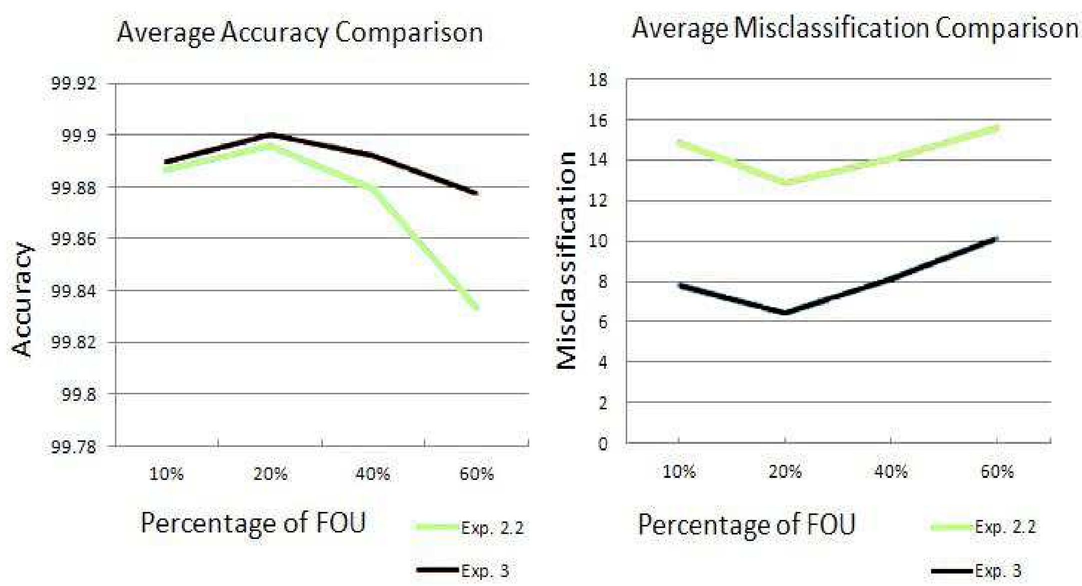


Figure 3.27: The experiment of “Crowded outdoor environment with snow”.



(a) (b)

Figure 3.28: Average misclassification and accuracy comparison in different frames for Experiment 2.2 multiple person in an outdoor snowy environment. (a) Average misclassification in different frames (b) Average accuracy in different frames



(a) (b)

Figure 3.29: Average misclassification and accuracy comparison in different FOU of IT2FLS for Experiment 2.2 multiple person in outdoor snowy environment and Experiment 3. crowded outdoor environment in snowy condition. (a) Average misclassification in different FOU (b) Average accuracy in different FOU.3

3.5 Discussion

In this chapter, we presented an interval type-2 fuzzy logic system for an improved silhouette extraction in dynamic real-world environments. Due to the huge complexity of real-life environments, the problem of detaching moving objects from human silhouette is complicated. To address this problem without a high computational complexity, we firstly use GMM to detect the foreground, then an interval type-2 fuzzy logic system is employed for moving objects detachment. We conducted several real-world experiments which have shown that the proposed interval type-2 fuzzy logic system is effective in detaching objects and the misclassification is greatly reduced compared to a similar type-1 fuzzy logic system while the interval type-2 fuzzy logic system also results in high accuracy for silhouette extraction compared to the type-1 fuzzy logic system. Moreover, experiments on different FOU and frames have been conducted to demonstrate the robustness of the proposed interval type-2 fuzzy logic system. Hence, by utilizing an interval type-2 fuzzy logic system, the proposed system obtains silhouette extraction with good robustness in relation to noise factors and existing uncertainties such as light condition changes, reflection of human body, and moving objects attached to the human silhouette, etc., in dynamic indoor/outdoor environments.

Chapter 4: A BB-BC based IT2FLS for the behaviour recognition on 2D video data in Intelligent Environments

4.1 Introduction

Recent years have witnessed a significant progress in the automation of human behaviour recognition using machine vision in order to realize intelligent environments which are capable of detecting users' actions and gestures so that the needed services can be provided automatically and instantly to users in such a way as to maximize user comfort and safety as well as minimizing energy. However, the majority of traditional human behaviour machine vision based recognition approaches rely on assumptions (such as known spatial locations and temporal segmentations) or computationally expensive approaches (such as sliding window search through a spatio-temporal volume). Hence, it is difficult for such methods to scale up and handle the high uncertainty levels and complexities available in real-world applications. In this chapter, we propose a novel fuzzy machine vision based framework for efficient humans' behaviour recognition. A model based feature set is utilised to extract visual feature cues including silhouette slices and movement speed from the human silhouette in video sequences which are analysed as inputs by the proposed fuzzy system. In this chapter, we employed fuzzy c-means clustering to acquire the membership functions of the proposed fuzzy system. The behaviour recognition was implemented via selecting the best candidate's behaviour category with the highest output degree as the recognised behaviour.

In this chapter, we will present several experiments which were performed on the publicly available Weizmann human action dataset to fairly compare them with the state-of-the-art algorithms. The experimental results demonstrate that the proposed optimization paradigm is effective in tuning the parameters of the membership functions and the rule base of the IT2FLSs to improve the recognition accuracy where the proposed IT2FLSs outperformed the Type-1 FLSs (T1FLSs) counterpart as well as other traditional non-fuzzy systems.

4.2 The Proposed Fuzzy based Human Behaviour Recognition

It is worth pointing out that categorizing human behaviour into one of several behaviour classes falls under the generic class of pattern recognition problem which aims to determine the mapping between behavioural feature space and action categories. Unfortunately, behaviour features of different subjects which are representative of the same action classes have a wide variance. In addition, the behaviour of a given subject conducting multiple instances of the same action category is not unique. In fact, there are intra- and inter- subject variations in behavioural characteristics which cause uncertainty in the behaviour recognition problem. Fuzzy logic system is an established field of research which handles uncertainties in complicated real world problems. In this chapter, we will employ fuzzy logic systems in order to handle the faced uncertainties associated with humans' behaviour recognition in intelligent spaces.

Figure 4.1 shows an overview of the proposed system where in the training stage a human silhouette is detected and extracted using our previous work based on an Interval Type-2 Fuzzy Logic System (IT2FLS) [Yao 2012]. After that, from the extracted silhouette, input feature vectors are computed for our fuzzy-based recognition method based on a model based feature set (see section 2.1) to describe the shape and motion characteristics. The fuzzy membership functions of the inputs to the fuzzy systems are then acquired via Fuzzy C-Means clustering. During the testing stage, we first tried to detect human subjects which were later tracked to extract the silhouette image based on which input shape-motion features are computed and used as input values for the fuzzy-based recognition system. In our fuzzy system, each membership function corresponds to a behaviour model while each output degree represents the likelihood between the behaviour in current frame and the trained behaviour model in the knowledge base. The behaviour in the current frame is then classified and recognised by selecting the candidate model which has the highest output degree. In the following subsection, we will present the various components of the proposed system.

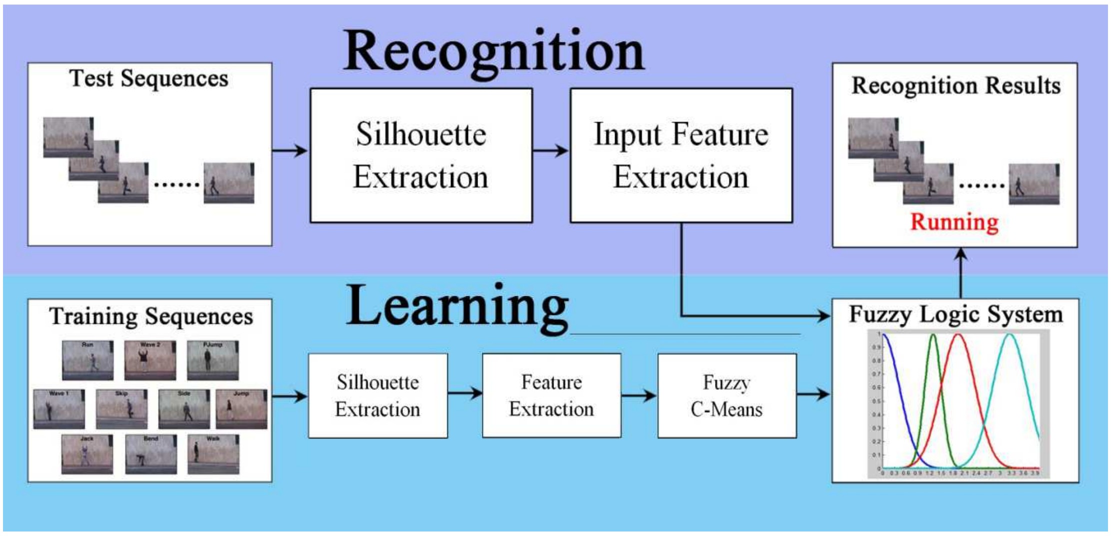


Figure 4.30: Overview of our proposed system.

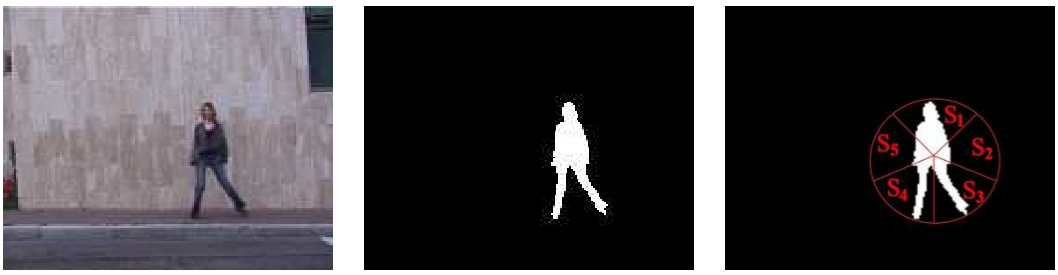
4.2.1 Silhouette Extraction

Accurate human silhouette segmentation from a video sequence is important and fundamental for advanced video procedures such as pedestrian detection, human activity detection and behaviour recognition. In order to obtain robust silhouette segmentation, the Gaussian Mixture Models (GMM) was proposed in [Katz 2003] which can extract the foreground images. However, it is unreasonable to simply consider GMM foreground as a human silhouette in real-world environments due to the noise factors including varying light conditions, reflections/shadows problems and moving objects attached to a human silhouette. To deal with these problems, a Type-1 Fuzzy Logic System (T1FLS) was proposed in [Chen 2006]. This T1FLS is capable of handling to an extent the uncertainties mentioned above, however, the extracted silhouette will be degraded due to misclassification. Hence, in [Yao 2012], we proposed an Interval Type-2 Fuzzy Logic System which was able to handle the high uncertainty levels present in real-world dynamic environments while also effectively reducing the misclassification of extracted silhouette. By utilising our proposed IT2FLS, the average accuracy for silhouette extraction is improved to a *99.94*% which is *8.26*% higher than the accuracy achieved by the T1FLS employed in [Chen 2006], meanwhile, the average misclassification of our proposed IT2FLS is reduced to *5.71* pixels which was *446.26* lower than the misclassification of the T1FLS in [Chen 2006].

4.2.2 Feature Representation

Our approach uses an efficient feature set [Vezzani 2010] with low computational complexity based on multi-feature model including movement speed and appearance shape. Based on the extracted silhouette image, as shown in Figure 4.2, the silhouette region is separated into five slices S_1, S_2, S_3, S_4, S_5 according to polar coordinates partitioning centred at the gravity centre $\{x_c(t), y_c(t)\}$. Ideally, the divided slices should be located at the areas of the head, the arms and the legs of the human silhouette. Suppose that we are working on frame t, and the human silhouette image of frame t is extracted from the proposed algorithm based on IT2FLS. Then, in the obtained silhouette image, as shown in Figure 6.c, the area of the entire human silhouette is denoted by the letter A_t while the areas of each silhouette slice are denoted by $\{A_t^i\}_{i=1\ldots5}$. Based on these values, the 7-dimensional feature set for our fuzzy-based recognition system is constructed, which is similar to the 17-dimensional the feature set in [Vezzani 2010] but which uses fewer input feature categories and a lower computational complexity. The seven input features are motion speed in horizontal direction (O^1), motion speed in vertical direction (O^2), area ratio of the head silhouette (O^3), area ratio of the right hand sillouette (O^4), area ratio of the right leg silhouette (O^5), area ratio of the left hand silhouette (O^6), and area ratio of the left leg silhouette (O^7). Thus the feature set contains both motion information (O^1, O^2) and shape description ($O^3 \ldots O^7$). The 7-dimensional feature set for the fuzzy system is obtained as follows:

$$O_t = \{O_t^1 \ldots O_t^7\} = \left\{ \begin{array}{l} O_t^1 = \sum_{i=0}^{2} | x_c(t-i) - x_c(t-i-1) | /3 \\ O_t^2 = \sum_{i=0}^{2} | y_c(t-i) - y_c(t-i-1) | /3 \\ O_t^{k+2} = \dfrac{A_t^k}{A_t}, k = 1 \ldots 5 \end{array} \right\} \tag{4.1}$$



(a) (b) (c)

Figure 4.31: (a) original image, (b) silhouette image extracted by our proposed method based on IT2FLS, (c) silhouette slice partitions.

4.2.3 The Proposed Type-1 Fuzzy System for Behaviour Recognition

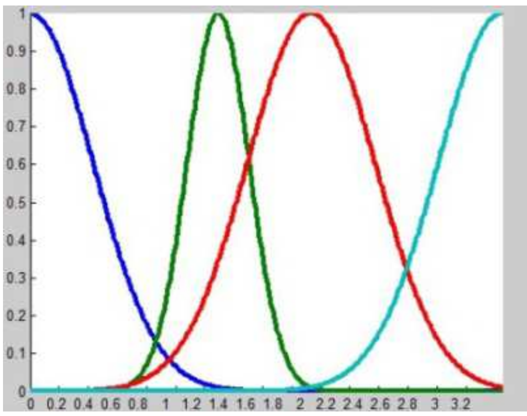
In our type-1 fuzzy system, the antecedents are seven linguistic variables which are: motion speed in horizontal direction (O^1), motion speed in vertical direction (O^2), area ratio of the head silhouette (O^3), area ratio of the right hand silhouette (O^4), area ratio of the right leg sillouette (O^5), area ratio of the left hand silhouette (O^6), and area ratio of the left leg silhouette (O^7). Each of these antecedents is represented by four fuzzy sets which are VERY LOW, 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 fuzzy Membership Function (MFs) shown in Figure 4.3 has been obtained via Fuzzy C Means (FCM)-based algorithm. In this FCM-based algorithm, the centres clustered by FCM will be used as the Means m_k of our Gaussian membership functions $(m_k, \sigma_k, x) = \exp\left(-\frac{1}{2}\left(\frac{x-m_k}{\sigma_k}\right)\right)$, where k = 1,…, p; p is the number of antecedents. And the variations σ_k were obtained by σ_k = $(m_k - m_{k-1})/3$, where k = 2,…, p; p is the number of antecedents. And for k = 1, σ_k = $(m_{k+1} - m_k)/3$

Suppose, we measure $\{O^1 \ldots O^7\}$ on silhouette images expressing the possibilities of the candidate behaviour classes: *running*, *walking*, *jumping-in-place*, *jumping-jack*, *jumping-forward*, *galloping-sideways*, *waving-two-hands*, *skipping*, *bending*, *waving-one-hand*. The mapping between measurement and behaviour classes is accomplished by fuzzy rules. In our system, the size of the rule based is 191, and the rule base is constructed via learning from the input/output data by experts in human behavioural analysis. One illustrative fuzzy rule from our rule base could be written as follows:

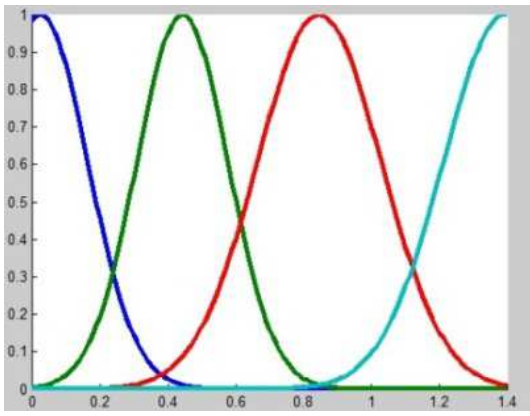
Rule *i*: IF (*horizontal-motion-speed* is HIGH) & (*head-silhouette-ratio* is MEDIUM) (*vertical-motion-speed* is VERY LOW) & (*leftHand-silhouette-ratio is* HIGH) *&* (*rightHand-silhouette-ratio is* LOW) *&* (*leftLeg-silhouette-ratio is* LOW) *&* (*rightLeg-silhouette-ratio is* LOW) THEN

(*running-possibility* is HIGH), (*otherBehaviours-possibility* is LOW)

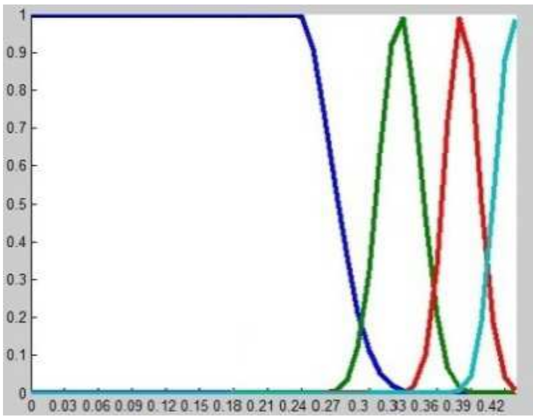
Each behaviour class uses the same output membership function that is shown in Figure 4.3h. In our system, we use product t-norm to represent the AND logical connective and the implication operation. The behaviour recognition is done via selecting the best candidate behaviour class with the highest firing strength as the recognised 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 behavioural similarity and cannot be distinguished effectively in the current frame.



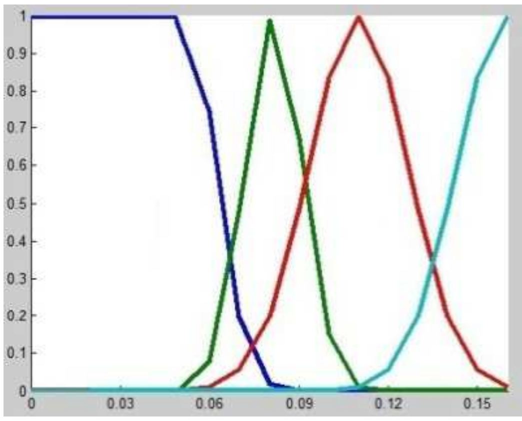
(a)



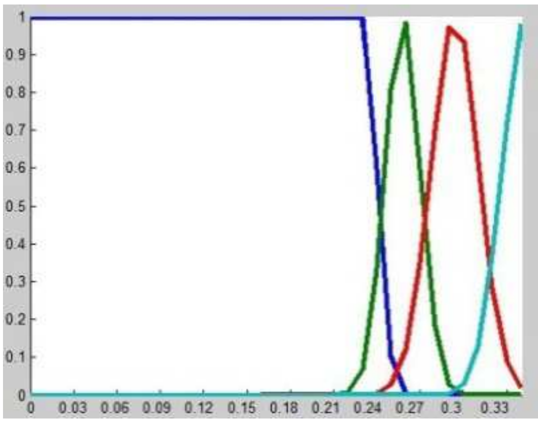
(b)



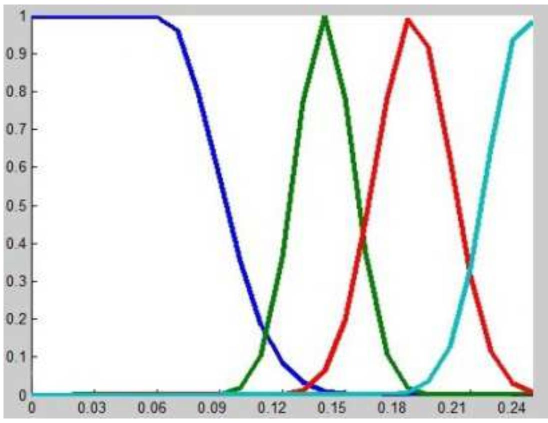
(c)



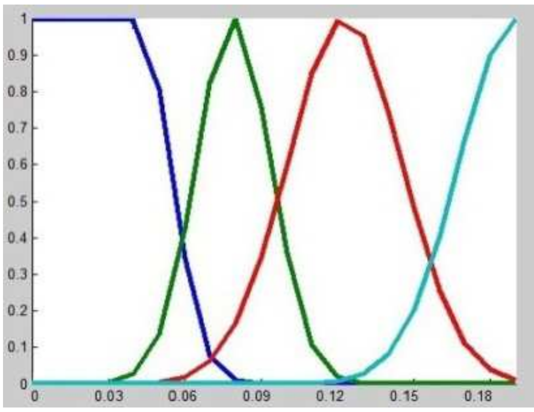
(d)



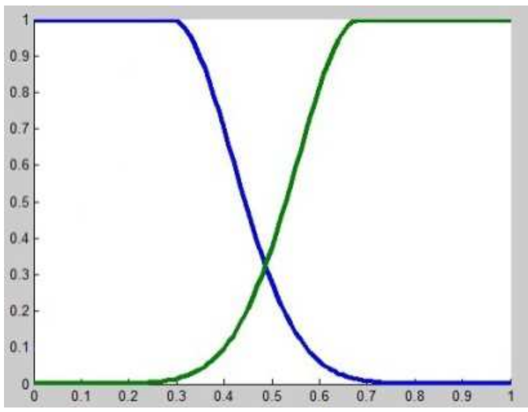
(e)



(f)



(g)



(h)

Figure 4.32: Membership functions from FCM for our fuzzy-based recognition system. a) MF for O^1, b) MF for O^2, c) MF for O^3, d) MF for O^4, e) MF for O^5, f) MF for O^6, g) MF for O^7, (h) MF for *Output.*

4.2.4 The Proposed Interval Type-2 Fuzzy System for Behaviour Recognition

In this subsection, we will present the manual design process of the IT2FLS which will be optimized by the proposed BB-BC algorithm presented in the next section. In the experiment section, we will 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 Figure 4.4, for example, the fuzzy set MEDIUM is represented by the Gaussian membership function whose mean and standard deviation are obtained from numerous measurements of the feature horizontal-motion-speed.

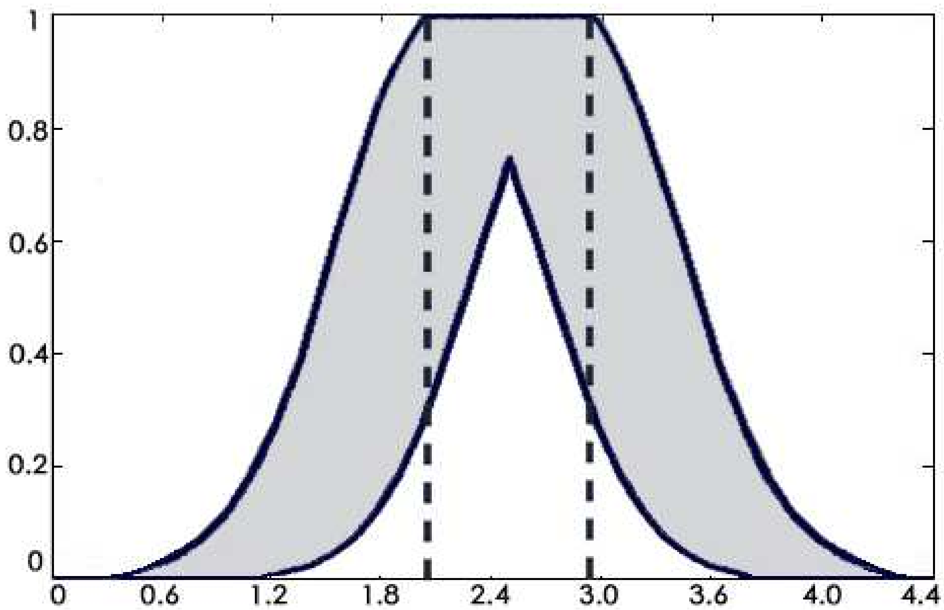


Figure 4.33: Gaussian MF of numerous instances of feature horizontal-motion-speed for the linguistic variable MEDIUM

To construct the type-2 MFs modelling the FOU, we transform the type-1 fuzzy sets (shown in Figure 4.4) 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 $[m_{k1}^{l}, m_{k2}^{l}]$ [Mendel 2001], i.e.,

$$u_{k}^{l}(x_k) = exp[-\frac{1}{2}\left(\frac{x_k - m_k^l}{\sigma_k^l}\right)],\ m_k^l \in [m_{k1}^{l}, m_{k2}^{l}] \qquad (4.2)$$

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

$$\bar{u}_{k}^{l}(x_k) = \begin{cases} N\left(m_{k1}^{l}, \sigma_k^l, x_k\right), & x_k < m_k^l \\ 1, & m_{k1}^{l} \leq x_k \leq m_{k2}^{l} \\ N(m_{k2}^{l}, \sigma_k^l, x_k), & x_k > m_{k2}^{l} \end{cases} \qquad (4.3)$$

And the lower membership function can be written as follows:

$$\underline{u}_{k}^{l}(x_k) = \begin{cases} N\left(m_{k2}^{l}, \sigma_k^l, x_k\right), & x_k \leq \frac{m_{k1}^{l} + m_{k2}^{l}}{2} \\ N\left(m_{k1}^{l}, \sigma_k^l, x_k\right), & x_k > \frac{m_{k1}^{l} + m_{k2}^{l}}{2} \end{cases} \qquad (4.4)$$

where

$$N\left(m_{k1}^{l}, \sigma_k^l, x_k\right) = exp\left(-\frac{1}{2}\left(\frac{x_k - m_{k1}^{l}}{\sigma_k^l}\right)\right) \qquad (4.5)$$

In our IT2FLS, in order to construct the type-2 MFs 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

$$m_{k2}^{l} = (1 + \alpha\%)\, m_{k1}^{l} \qquad (4.6)$$

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 rule base for the IT2 FLS and the T1FLS.

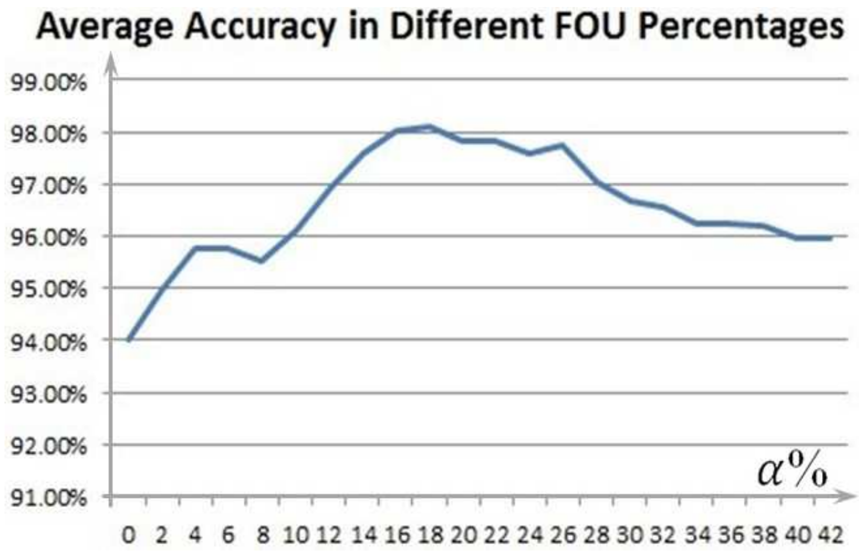


Figure 4.34: Average accuracy comparison in different FOU of IT2FLS. The horizontal axis represents α% used in the system while the vertical axis is the average recognition accuracy.

The uncertainty factor α was chosen in order to maximise the achieved accuracy. Figure 4.5 shows the accuracy plotted against the various sizes of FOUs. It is obvious that the best IT2FLS results in the highest average accuracy which is obtained by using the uncertainty factor of α=18%. As can be seen in Figure 4.5, when the FOU percentage is α=0, the type-2 system is degraded to the type-1 system, and when the FOU percentage increases, the accuracy of the type-2 system is higher than the type-1 system proving that the IT2FLSs (with any FOU) are more reliable in handling the uncertainties in the given environments.

Note that the α=18% is the same for all the fuzzy sets for the various inputs. We will employ the type-2 fuzzy sets α=18% as the initial population for the BB-BC algorithm which will optimize the values of α which might be different for the various input and output type-2 fuzzy sets.

4.2.5 The Proposed Optimization Method for the IT2FLS

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 in the direction of having higher accuracy.

4.3 Big Bang-Big Crunch (BB-BC) Optimization

A. 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 (4.6), 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, as in the MFs of proposed system, four type-2 fuzzy sets are utilized for modelling each of the *7* features linguistic terms including VERY-LOW, LOW, MEDIUM and HIGH, the total number of the parameters for the input type-2 MFs is 4×7=28. In a similar manner, parameters for the output MF are also encoded which 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 Figure 4.6.

α_1^1	α_2^1	α_3^1	α_4^1	$\cdots$	α_1^7	α_2^7	α_3^7	α_4^7	α_L^{Out}	α_H^{Out}

Figure 4.35: 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 shown in Figure 4.6 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:

$$f^i = (1 - Accuracy^i) \tag{4.7}$$

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

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

In a similar fashion of 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 Figure 4.7.

o_1^1	o_2^1	$\cdots$	o_7^1	o_{out}^1	$\cdots$	o_1^R	o_2^R	$\cdots$	o_7^R	o_{out}^R

Figure 4.36: The population representation for the parameters of type-2 MFs

As shown in Figure 4.7, o_k^r are the antecedents and o_{out}^r is the consequent of each rule, respectively, where *k* = 1,…, *p*, *p* is the number of antecedents; *r* = 1,…, *R*, *R* is the number of the rules to be tuned. In this study, the rule base constructed by the experts is employed as the initial generation of candidates. After that, the rule base can be tuned by BB-BC using the cost function depicted in Equation (4.7).

4.4 Experiment Results

We tested the proposed system on the widely used benchmark datasets for humans' behaviour recognition, the Weizmann human action dataset [Blank 2005]. The Weizmann actions dataset consists of *5687* frames and *10* different categories of behaviour classes (as shown in Figure 4.8): running, walking, jumping-in-place-on-two-legs (pjump), jumping-forward (jump), bending, jumping-jack (jack), galloping-sideways (side), skipping, waving-two-hands (wave2), waving-one-hand (wave1). Video sequences in this dataset are captured with a stationary camera and a simple background. However, it provides a good experiment environment to investigate the recognition accuracy of the proposed method when the number of behaviour categories is large. Example frames of the behaviour categories are shown in Figure 4.8. Each behaviour category is performed once or sometimes twice per video by nine different people (subjects) resulting in *93* video sequences in total.

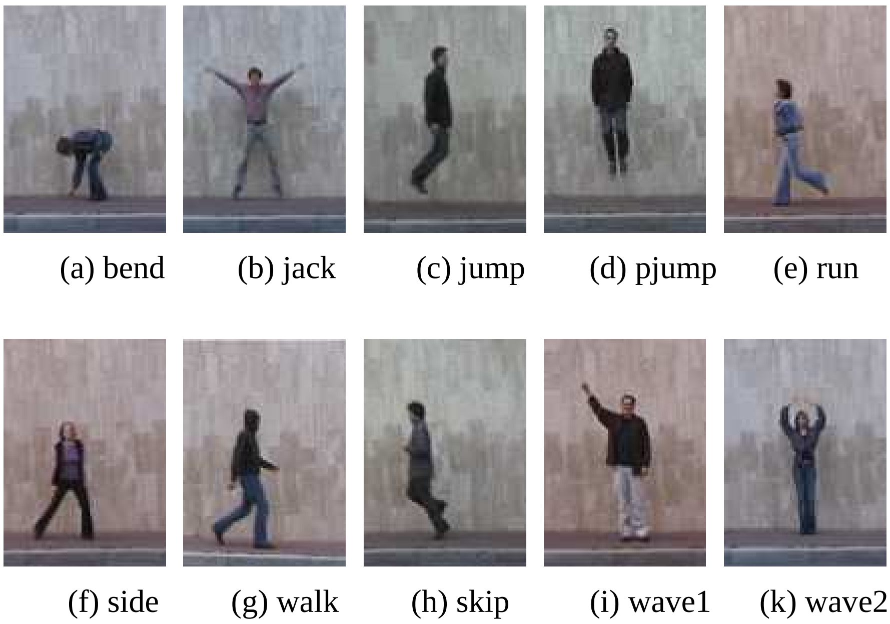
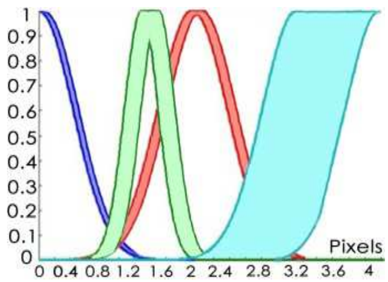


Figure 4.37: Weizmann dataset for human behaviour recognition. a) bending, b) jumping-jack , c) jumping-forward, d) jumping-in-place-on-two-legs, e) running, f) galloping-sideways, g) walking, h) skipping, i) waving-one-hand, k) waving-two-hands.

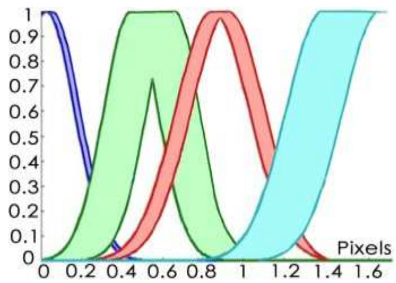
The type-2 membership functions obtained and optimized by BB-BC are shown in Figure 20. By utilizing BB-BC to optimize the MFs and rule base, the accuracy of per-frame recognition of our IT2FLS is improved to 100%. Table 4.2

demonstrates that the BB-BC optimization improves the performance of the T1FLS and the IT2FLS where the per-frame accuracy is improved when using BB-BC optimization for either the T1FLS or the IT2FLS. As shown in Table 4.2, our IT2FLSs-based method with BB-BC optimization achieves 1.88% higher average per-frame accuracy than the IT2FLS-based system without BB-BC optimization (using manual rule base and α=18%).

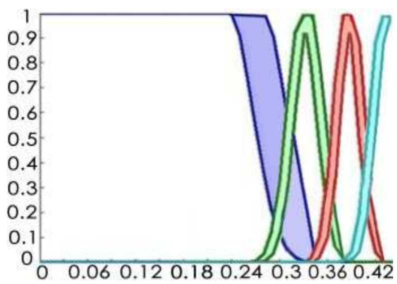
Also note that the IT2FLSs-based system outperforms its counterpart T1FLSs-based recognition system with the same rule base (either manually designed or optimized by BB-BC). As shown in Table 4.2, our manually designed IT2FLSs-based method (using manual rule base and α=18%) achieves 4.09% higher average per-frame accuracy than the manually designed T1FLS. In addition, the BB-BC optimized IT2FLS achieves 4.55% higher average per-frame accuracy than the BB-BC optimized T1FLS.



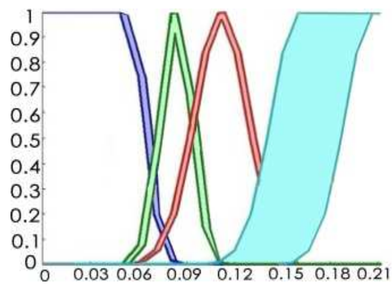
(a)



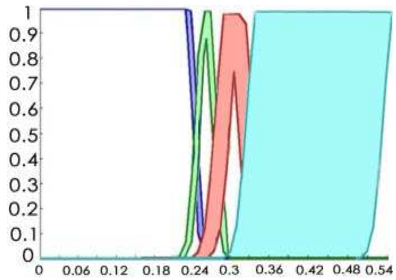
(b)



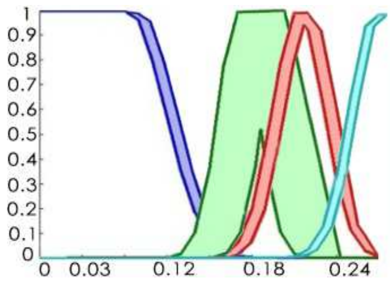
(c)



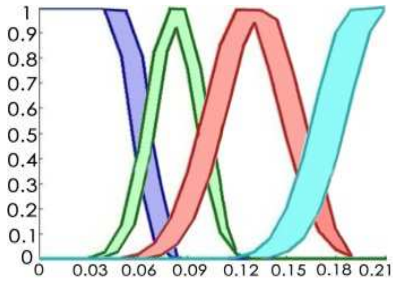
(d)



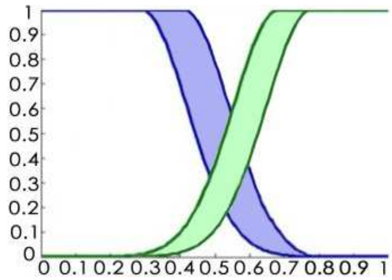
(e)



(f)



(g)



(h)

Figure 4.38: Type-2 membership functions optimized by using BB-BC. a) Type-2 MF for O^1, b) Type-2 MF for O^2, c) Type-2 MF for O^3, d) Type-2 MF for O^4, e) Type-2 MF for O^5, f) Type-2 MF for O^6, g) Type-2 MF for O^7, h) Type-2 MF for Output.

Similar to [Vezzani 2010], our experiments are performed by leave-one-out cross validation as suggested by Scovanner et al. [Scovanner 2007]. During the testing stage, our model was evaluated for per-frame and per-video recognition. Specifically, per-frame recognition means performing the analysis algorithm on each frame and then obtaining a recognition result for each individual frame, while per-frame recognition denotes achieving a global recognition result for the entire video sequence. The average accuracy of our method and comparison with other traditional non-fuzzy based algorithms are reported in Table 4.1.

Our type-2 fuzzy logic system also outperforms the traditional non-fuzzy based recognition method which used hidden Markov models (HMM) [Vezzani 2010]. In order to conduct a fair comparison with the traditional HMM-based method [Vezzani 2010], our IT2FLSs-based system utilizes similar input features to the model proposed in [Vezzani 2010] which employed 7 feature categories out of 17 feature categories from the feature set in [Vezzani 2010]. As shown in Table 4.1, our IT2FLSs-based method with BB-BC optimization achieves 13.3% higher average per-frame accuracy than the HMM-based algorithm.

By using the proposed optimization method, our system outperformed the per-frame recognition accuracy of other state-of-the-art method based on hidden conditional random field (hCRF) [Wang 2008] by 9.71% and our method has also outperformed the k-NN-based algorithm-based algorithm [Niebles 2007] by 45% recognition accuracy. The per-video accuracy of the proposed method is 100% outperforming the traditional non-fuzzy approaches including hCRF-based method [Wang 2008], SVM-based approach [Jhuang 2007] and k-NN-based algorithm [Niebles 2007] by 2.78%, 1.20% and 27.20%, respectively. In our experiment,

Methods	per-frame	per-video	frame rate
IT2FLSs with BB-BC	100%	100%	30 f/s
IT2FLSs without BB-BC	98.12%	100%	30 f/s
T1FLSs with BB-BC	95.45%	100%	30 f/s
T1FLSs without BB-BC	94.03%	100%	30 f/s
Jiang et al. [Jiang 2012]	100%	100%	1.06 f/s
Vezzani et al. [Vezzani 2010]	86.70%	N/A	15 f/s
Wang & Mori [Wang 2008]	90.29%	97.22%	N/A
Jhuang et al. [Jhuang 2007]	N/A	98.80%	0.43f/s
Niebles & Fei-Fei [Niebles 2007]	55.00%	72.80%	N/A

Table 4. 1: Comparison of Overall Average Recognition Accuracy with Previous Traditional Non-fuzzy Methods on the Weizmann Dataset.

In our experiments of this chapter, nine human subjects were involved in ninety three video sequences. And we performed nine sessions run to analyse to the results. In In our experiments of this chapter, instructed behaviour was used in all of the tests.

Our IT2FLSs-based approaches are computationally efficient. The proposed IT2FLSs-based recognition system (including the background subtraction and updating, human silhouette extraction, multi-target tracking, feature extraction and behaviour recognition) works processing 30 frames per second in real time. As shown in Table 4.1, our method improves by 100% the computation performance when compared with the conventional HMM-based algorithm [Vezzani 2010] which has a frame rate of 15 frames per second. We outperformed the computation speed of SVM-based method in [Jhuang 2007] by 6876.74% where the SVM approach can only process 0.43 frames per second. And similarly, the proposed IT2FLSs-based system has also outperformed the computation performance of tree-based approach in [Jiang 2012] by *2730.18*%

4.5 Discussion

This chapter presented a computationally efficient system based on interval type-2 fuzzy logic systems for the automatic recognition of human behaviour using machine vision for applications in intelligent environments. It is hoped that the proposed method will be an enabling step towards the realization of ambient intelligent environments which can automatically detect the human behaviour and adapt the user environment accordingly. In our system, the original images are first captured from the input video sequences and the extracted human silhouette is generated using our proposed method based on interval type-2 fuzzy logic system. After that, the input features are computed from the extracted silhouette images using a seven-dimensional model-based feature set including motion information and shape descriptors. Finally, human behaviour is recognized based on the input feature set by using the proposed IT2FLSs-based recognition method. The experiment results demonstrate the superior performance of the proposed BB-BC based approach for optimizing the membership functions and rule base of the IT2FLS which outperformed the equivalent T1FLS and the state-of-the-art non-fuzzy methods regarding recognition accuracy and analysis performance.

Chapter 5: A BB-BC based IT2FLS for the Event Detection and Summarisation using 3D sensor in real-world Ambient Assisted Living Environments

5.1 Introduction

Video monitoring can provide vital context awareness information from indoor intelligent environments where privacy is not a constraint. However, there is a need to develop linguistic summarization tools which are capable of summarizing in layman's terms the information which is of interest in long video sequences. The key module which can enable the linguistic summarization of video monitoring is human activity/behaviour recognition. Human behaviour recognition is an important yet challenging task due to the behaviour uncertainty, activity ambiguity, and uncertain factors such as position, orientation and speed, among others. 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. In this chapter, we will 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 along with other traditional non-fuzzy based systems. Based on the recognition results, higher-level applications will be presented including video linguistic summarizations event searching and activity retrieval/playback.

5.2 Overview of the Employed Hardware of RGB-D Sensor

The Kinect has become the most popular RGB-D sensor in recent years. Most of the other RGB-D sensors, such as ASUS Xtion and PrimeSense Capri as shown in Figure 5.1a and 5.1b, normally use the PS1080 hardware design and chip from PrimeSense, which was bought by Apple in 2013.

The original Kinect v1 camera, as demonstrated in Figure 5.1c, was first introduced in 2010 and was mainly used to capture users' body movements and motions for interacting with the program, but it was rapidly repurposed to be utilized in a diverse array of novel applications from healthcare to robotics. It has been repurposed in the field of intelligent environments and robotics as an affordable but robust replacement for various types of wearable sensors, expensive distance sensors and conventional 2D cameras. It has been successfully used in various applications including object tracking and recognition [Han 2013] as well as 3D indoor mapping and human activity analysis [Yao 2014]. However, the structured-light technology of Kinect v1 limited the usage of its depth camera in outdoor environments where it could not sense minor objects, and it had depth resolutions (320×240) and field of view (57°×43°) that were too low to satisfy the needs and requirements of some of the real-world application scenarios. By contrast, the new generation Kinect v2, as presented in Figure 5.1d, was completely improved to employ time-of-flight range sensing where the infrared camera ejects strobe infrared light into the scene, and calculates the time length for the bursts of light to return to each pixel. In this way, its infrared camera can produce high-resolution (512×424) depth images at a field of view of 70°×60°. At the same time, Kinect v2 produces high-resolution (up to 1920×1080) colour images at a field of view of 84°×53° using a build-in colour camera which performs as well as a regular high-definition (HD) CCTV camera. One of the extra merits of the Kinect v2 is its low price at about £130 as well as its convenient software development kit (SDK) which can return various robust features such as 3D skeleton data for rapid development and research. The limitations of the Kinect v2 sensor are: (1) if the human subject is around the boundary of the view, the quality of skeleton detection would degrade and further leads to poor analysis result; (2) if the distance of the human subject farther than 4.3 meter away from sensor, the quality of skeleton detection would become poor. (3) Within the field of view 70°×60° and the distance range 4.5 meter, if there are more than 4 people in the scene, it would be crowded and lead to serious occlusion problem which causes bad recognition performance.

(a) **(b)**

(c) **(d)**

Figure 5.39: Main RGB-D sensors. a) ASUS Xtion, b) PrimeSense Capri, c) Kinect v1, d) Kinect v2.

For most of the user-oriented applications in intelligent environments and healthcare, the features of the user's posture, especially skeleton data, make up the core information since the skeleton data describes the skeleton's joint positions and orientations of the user in the scene. There are several skeleton trackers available including Kinect skeleton tracker, Open Natural Interaction (OpenNI/NiTE) skeleton tracker, and Point Cloud Library (PCL) skeleton tracker [Rusu 2011]. For the Kinect skeleton tracker, a random decision forest-based method [Shotton 2013] is used in Kinect v1 to robustly extract the 20 joints from one subject. In the SDK of Kinect v2, the skeleton tracker is improved and can robustly extract up to 25 3D joints (as shown in Figure 5.2) from a single user (with new joints for hands and neck, etc.). It also handles the occlusion problem of different users and supports up to 6 users at a scene at the same time. The effective sensing range of the Kinect skeleton tracker is from 0.4 meters to 4.5 meters. In the PrimeSense's OpenNI, a skeleton tracker was provided and can extract the positions of 15 joints from a single user. However, in our experiment tests, it is not as accurate as the Kinect skeleton tracker but it supports backward skeleton extraction. For the PCL skeleton tracker, 15 joints can be analysed from a subject but this module requires a video card supporting nVidia CUDA and the PCL skeleton tracker is less robust than the other trackers according to our experiment tests. Therefore we use the Kinect skeleton tracker in our system.

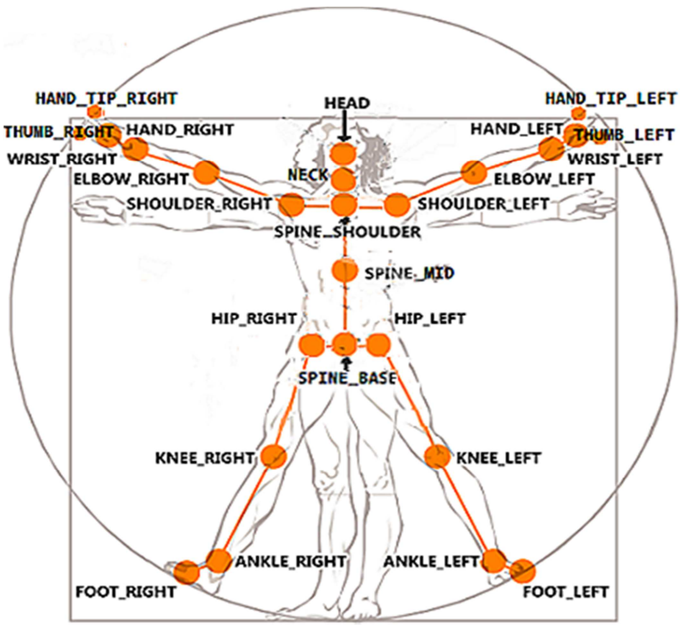


Figure 5.40: 3D Skeleton and joints of Kinect v2.

5.3 The Proposed BB-BC based Interval Type-2 Fuzzy Logic System for the Event Detection and Linguistic Summarization of Video Monitoring

Figure 5.3 provides an overview of our proposed system. There are two phases in the system; these are the learning phase and the recognition phase. During the learning phase, the training data for each behaviour category are collected from the real-time Kinect data captured from the subjects in different circumstances and situations. Then behaviour feature vectors based on the distance and angle feature information are computed and extracted from collected Kinect data so as to model the motion characteristics. From the results of the features extraction, the type-1 fuzzy Membership Functions (T1MFs) of the fuzzy systems are recognized/known/discovered via Fuzzy C-Means Clustering (FCM) [Pal 1995]. After that, the type-2 fuzzy MFs are 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). Then, with the acquired membership functions, the rule base of the type-2 fuzzy system is constructed automatically from the input feature vectors. Finally, a method based on the BB-BC algorithm is used to optimize the parameters of the IT2FLS which will be employed to recognize the behaviour and activity in the recognition phase.

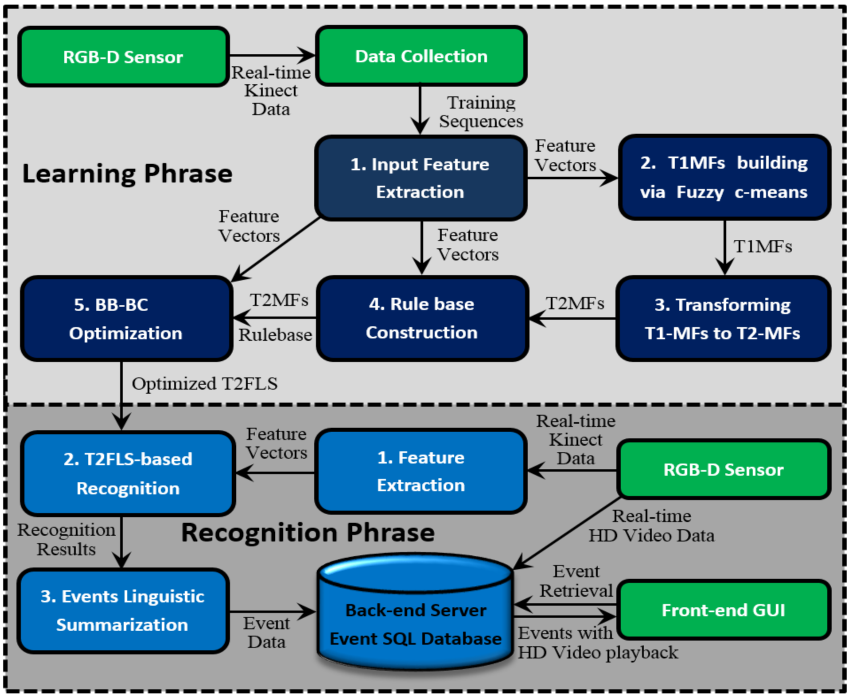


Figure 5.41: Overview of our proposed system.

It should be pointed that we generated initial fuzzy sets and rules for the FLSs which we then optimized via the BB-BC approach as such initial fuzzy sets and rules provided a good starting point for the BB-BC to converge quickly to the optimal position. If we had started from random fuzzy sets and rules, the BB-BC would have taken very long time to converge to optimal values.

During the recognition phase, the real-time Kinect data and HD video data are continuously captured by the RGB-D sensor monitoring the scene. From the real-time Kinect data, behaviour 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. Once important events are detected by the optimized IT2FLS, a linguistic summarization is performed using the key information such as the output action category, the starting time and ending time of the event, the user's number and identification, and the relevant HD video data and video descriptions. After that, the summarized event data is efficiently stored in our back-end server of event SQL database from where the users can access locally or remotely by using our front-end Graphical User Interface (GUI) system and perform event searching, retrieval and playback. The details of the employed phases are discussed in the following subsections.

5.4 Learning Phase

5.4.1 Fuzzy c-means

The Fuzzy c-mean (FCM) algorithm developed by Dunn [Dunn 1973] and later improved by Bezdek [Pal 1995] is an unsupervised clustering method used to classify the unlabelled data by minimizing an objective function. The FCM uses fuzzy partitioning such that each data point belongs to a cluster to a certain degree modelled by a membership degree in the range [0, 1], which indicates the strength of the association between that data point and a particular cluster centroid. Let $X = \{x_1, x_2, \dots, x_N\}$ be a set of given data points and $V = \{v_1, v_2, \dots, v_N\}$ be a set of cluster centres. The idea of the FCM is to conduct a partition of the N data points into C clusters based on minimization of the following objective function:

$$J(X; U, V) = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - v_j \right\|^2 \quad (5.1)$$

where m is used to adjust the weighting effect of membership values, $\| \cdot \|$ is the Euclidean norm modelling the similarity between the data point and the centre, and $U = (u_{ij})_{C \times N}$ is a fuzzy partition matrix subject to:

$$\sum_{i=1}^{C} u_{ij} = 1, \ \forall j = 1, \dots, N \quad (5.2)$$

and

$$u_{ij} \in [0, 1], \ \forall i = 1, \dots, C, \ \forall j = 1, \dots, N \quad (5.3)$$

where u_{ij} is the membership degree of point x_i to the cluster *j*. The FCM is performed via an iterative procedure with the Equation (5.1) updating u_{ij} and c_j.

In this chapter, the FCM is used to compute the clusters of each feature to generate the type-1 fuzzy membership functions for the fuzzy-based recognition system. The optimization procedure of FCM can be summarized in the following steps:

Step 1: Set the iteration terminating threshold ε to a small positive number in the range [0, 1], the weighting exponent$m$, and the number of clusters *C* (in our system, ε is set to be *0.0005*, *m* is initialized by using small positive random numbers ranging in [0, 1] and *C* is set to be *3* representing the fuzzy sets *LOW*, *MEDIUM*, *HIGH*) and set the number of iteration *t = 0*.

Step 2: Increase the number of iteration *t* by *1*

Step 3: Calculate the cluster centres by using the following equation:

$$v_i^{(t)} = \frac{\sum_{j=1}^{N} (u_{ij}^{(t-1)})^m x_j}{\sum_{j=1}^{N} (u_{ij}^{(t-1)})^m}, \ \forall i = 1, \dots, C \quad (5.4)$$

Step 4: Compute all the u_{ij} using the following equation to update the fuzzy partition matrix by the newly obtained u_{ij}

$$u_{ij}^{(t)} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_j - v_i^{(t)}\right\|}{\left\|x_j - v_k^{(t)}\right\|}\right)^{\frac{2}{m-1}}}, \quad \forall i = 1, \dots, C, \ \forall j = 1, \dots, N \tag{5.5}$$

Step 5: Check if $\left\|U^{(t)} - U^{(t-1)}\right\|^2 < \varepsilon$ then stop; otherwise go to Step 2.

These steps help identify the centre of each type-1 fuzzy set and the associated membership distribution. We will repeat the above steps for each input and output variable so as to extract their type-1 fuzzy sets membership functions.

5.4.2 Feature Extraction

5.4.2.1 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. For any three different 3D points, P_1, P_2, and P_3, an angle feature θ is defined by these three 3D joints P_1, P_2 and P_3 at a time instant. The angle θ is obtained by calculating the angle between the vectors $\overrightarrow{P_1P_2}$, and $\overrightarrow{P_2P_3}$ based on the following equation:

$$\theta = cos^{-1}\left(\frac{\overrightarrow{P_1P_2} \times \overrightarrow{P_2P_3}}{\left|\overrightarrow{P_1P_2}\right| \left|\overrightarrow{P_2P_3}\right|}\right) \tag{5.6}$$

5.4.2.2 Joint-position Feature Representation

In order to model the local "depth appearance" for the joints, the joint positions are computed to represent the motion of the skeleton. For distance, between joint *i* and joint *j*, the arc-length distance is calculated:

$$D_{ij} = \|P_i - P_j\| \quad (5.7)$$

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

5.4.2.3 Posture Representation

To perform efficient behaviour recognition, an appropriate posture representation is essential to model the gesture characteristics. In this work, we use Kinect v2 to extract the 3D skeleton data which comprises 3D joints, which are shown in Figure 5.4. After that, based on the 3D joints obtained, we compute the posture feature using the joint vectors as shown in Figure 5.4. In the applications of AAL environments, the main focus is to understand the users' daily activities and regular behaviour so as to create an ambient context awareness such that ambient assisted services can be provided to the users in the living environments. Therefore, in our current application of scenarios of ambient assisted living environments, we recognize and summarize the following kinds of behaviour: *drinking/eating, sitting, standing, walking, running, and lying/falling down* to provide different ambient assisted services. For example, if an elderly person falls down our system will send a warning message to the nearby caregivers or other people who can help. We also summarize the frequency of the drinking activity to ensure that the user drinks enough water throughout the day to avoid dehydration. By a daily summarization of the sitting and lying duration and frequency, healthcare advice would be provided if the user remains inactive/active most of the time. The detection results of running demonstrate a potential emergency happening. From the detection results of standing and walking, our system obtains the location and trajectory of the subject so that services, such as wandering prevention, can be provided to dementia patients. Also, the risk of falling down can be reduced by analysing the pattern of standing and walking. Furthermore, cognitive rehabilitation services can be provided to help the elderly with dementia by summarizing this series of daily activities. To achieve the robust recognition and summarization of the behaviour in AAL environments, we use the angles and distance of these joint vectors as the input features which are highly relevant when modelling the target behaviours in AAL environments. The identified behaviour is extendable to enlarge the recognition range of the target behaviour by adding the needed joints.

Since most kinds of daily behaviour (e.g., drinking, eating, waving hands, taking pills) are related to the upper body, in order to recognize behaviour and activity we focus in this work on the following joints: *spine base* (P_{sb}), *spine shoulder* (P_{ss}), *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 easily be added and utilized in 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 $\overrightarrow{P_{ss}P_{el}}$, $\overrightarrow{P_{ss}P_{hl}}$ modelling the left arm, and $\overrightarrow{P_{sc}P_{er}}$, $\overrightarrow{P_{sc}P_{er}}$ modelling the right arm.

(2) Angle features of the left arm θ_{al} can be obtained by calculating the angle between vectors $\overrightarrow{P_{ss}P_{el}}$, $\overrightarrow{P_{ss}P_{hl}}$ based on Equation (5.6). Similarly, angle features of the right arm θ_{ar} can be computed by applying the same process on $\overrightarrow{P_{ss}P_{er}}$, $\overrightarrow{P_{ss}P_{hr}}$.

(3) Based on Equation (5.7), position feature D_{hl}, D_{hr} of the vectors $\overrightarrow{P_{ss}P_{hl}}$, $\overrightarrow{P_{ss}P_{hr}}$ can be obtained. In order to recognize activities, the status (3D position and angle) of the spine of the human subject is modelled in a way that is invariant to orientation and position.

(4) Compute the vector $\overrightarrow{P_{ss}P_{sb}}$, modelling the entire spine of the subject, and $\overrightarrow{P_{ss}P_{kl}}$, $\overrightarrow{P_{ss}P_{kr}}$ modelling the left knee and right knee. Compute the angle θ_{kl} between $\overrightarrow{P_{ss}P_{sb}}$ and $\overrightarrow{P_{ss}P_{kl}}$ by using Equation (5.6). Similarly, the angle θ_{kr} can be obtained by applying Equation (5.6) on the vectors $\overrightarrow{P_{ss}P_{sb}}$ and $\overrightarrow{P_{ss}P_{kr}}$. Then, the bending angle θ_b of the body can be modelled, which is mainly used for analysing the *sitting* activity

$$\theta_b = max(\theta_{kl}, \theta_{kr}) \tag{5.8}$$

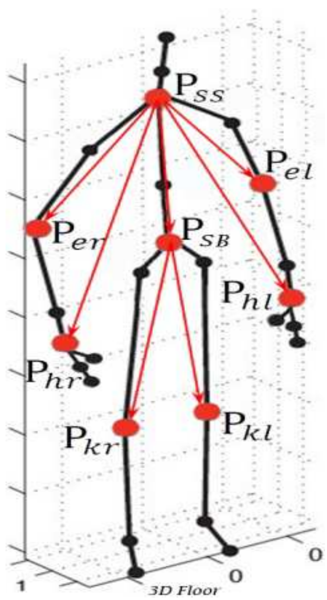


Figure 5.42: 3D feature vectors based on the Kinect v2 skeleton model

(5) In order to recognize the *lying/falling down* activity, we compute the distance D_f between the 3D coordinates *Spine Base* P_{sb} to the 3D Plane of the *floor* in the vertical direction.

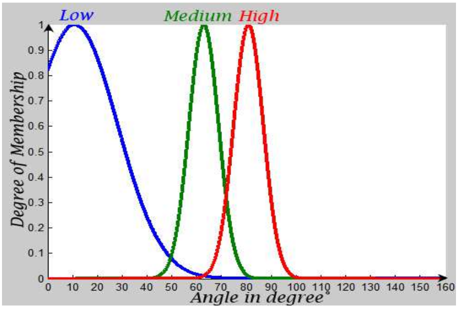
(6) We compute the movement speed of the human by analysing P_{sb}^{i-1} and P_{sb}^{i} which are the positions of the joint P_{sb} in two successive frames, frame *i-1* and frame *i*. The speed D_{sb} can be obtained by applying Equation (5.7) on P_{sb}^{i-1} and P_{sb}^{i}. The movement speed D_{sb} is mainly utilized for analyzing the common activities: *falling down*, *sitting*, *standing*, *walking*, and *running*.

For each tracked subject at a certain frame, the motion feature vector is obtained:

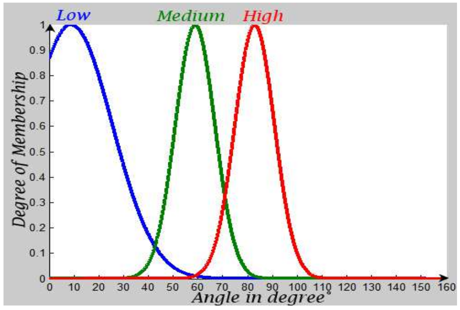
$$M = (\theta_{al}, \theta_{ar}, D_{hl}, D_{hr}, \theta_b, D_f, D_{sb}) \tag{5.9}$$

For simplicity, we also denote each feature in *M* using the following format:

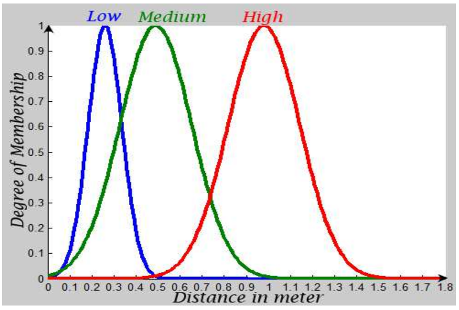
$$M = (m_1, m_2, m_3, m_4, m_5, m_6, m_7) \tag{5.10}$$



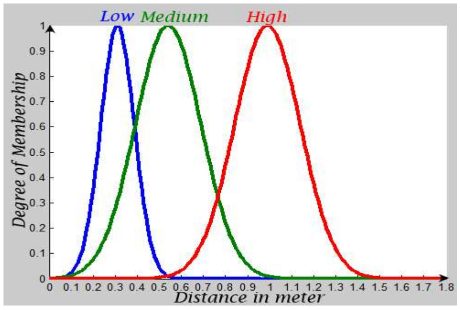
(a)



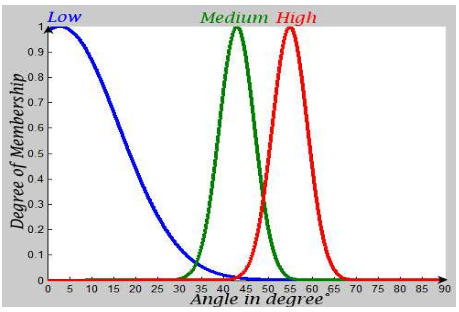
(b)



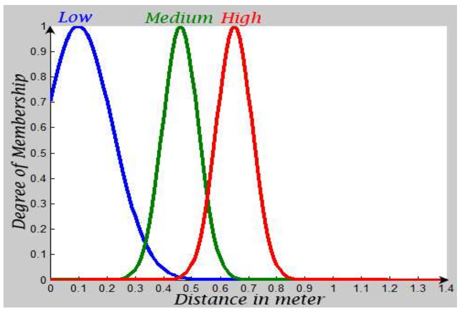
(c)



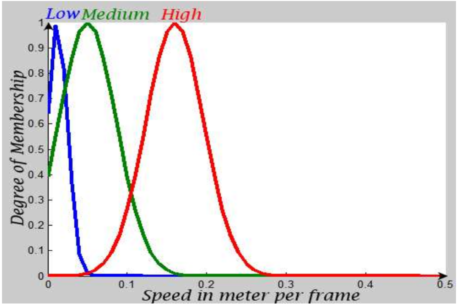
(d)



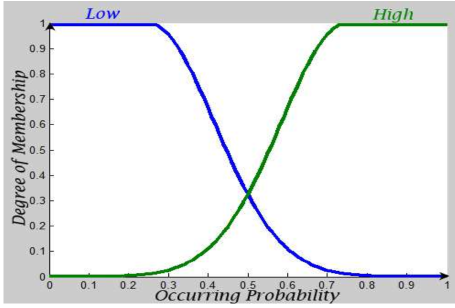
(e)



(f)



(g)



(h)

Figure 5.43: Type-1 membership functions constructed by using FCM, (a) Type-1 MF for m_1 (b) Type-1 MF for m_2 (c) Type-1 MF for m_3 (d) Type-1 MF for m_4 (e) Type-1 MF for m_5 (f) Type-1 MF for m_6 (g) Type-1 MF for m_7(h) Type-1 MF for the Outputs

As we can see, our system is a general framework for behaviour recognition, which can be easily extended to recognize more behaviour types by adding more relevant joints into the feature calculation.

5.4.2.4 Occlusion problems and Tracking State Reliability

For most available 3D motion capture devices in the market, the hardware system provides the level of the tracking reliability of the 3D joints. Kinect also returns to the tracking status to indicate if a 3D joint is *tracked* robustly, or *inferred* according to the neighbouring joints, or *not-tracked* when the joint is completely invisible. The 3D joints, which are occluded, belong to the *inferred* or *not-tracked* part. In our experiments, we found out that both *inferred* and *not-tracked* joints are unusually unreliable and cause misclassifications. Thus, to solve the occlusion problem and increase reliability, we only perform recognition when the tracking status of the essential parts related to our algorithm is *tracked.* In this way we can avoid misclassifications.

5.4.2.5 Transforming Type-1 Membership Functions to Interval Type-2 Membership Functions

In this subsection, we will present the initial design process of the IT2FLS, which will be further optimized by the proposed BB-BC algorithm presented in the next subsection. Figure 5.6 shows the type-1 fuzzy sets which were extracted via FCM.

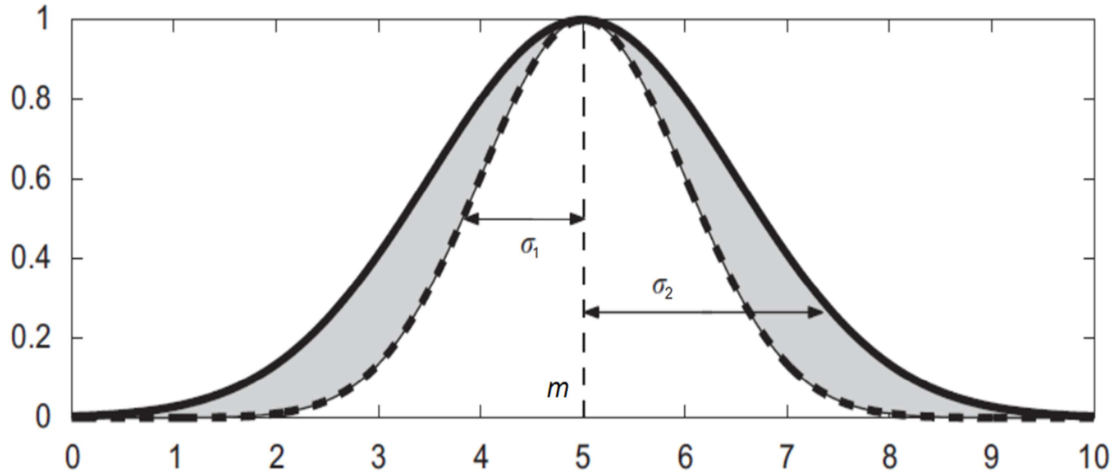


Figure 5.44: Example of the type-2 fuzzy membership function of the Gaussian membership function with uncertain standard deviation σ. The shaded region is the Footprint of Uncertainty(FOU). The thick solid and dashed lines denote the lower and the upper membership functions [Mendel 2001]

In order to construct the initial type-2 MFs modelling the FOU, we transform the type-1 fuzzy sets to the interval type-2 fuzzy sets with certain mean (m) and uncertain standard deviation σ $[\sigma_{k1}^{l}, \sigma_{k2}^{l}]$ [Hagras 2004], [Mendel 2001], i.e.,

$$\mu_k^l(x_k) = \exp[-\frac{1}{2}\left(\frac{x_k - m_k^l}{\sigma_k^l}\right)] \,, \ \sigma_k^l \in [\sigma_{k1}^l, \sigma_{k2}^l] \tag{5.11}$$

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

$$\bar{\mu}_k^l(x_k) = N(m_k^l, \sigma_{k2}^l, x_k) \tag{5.12}$$

The lower membership function can be written as follows:

$$\underline{\mu}_k^l(x_k) = N(m_k^l, \sigma_{k1}^l, x_k) \quad (5.13)$$

where

$$N\left(m_k^l, \sigma_k^l, x_k\right) = \exp\left(-\frac{1}{2}\left(\frac{x_k - m_k^l}{\sigma_k^l}\right)\right) \quad (5.14)$$

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

$$\sigma_{k2}^l = (1 + \alpha\%)\, \sigma_{k1}^l \quad (5.15)$$

where m_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 IT2FLS and the T1FLS.

5.4.2.6 Initialize Rule base construction from the raw data

In this chapter, we use the Wang-Mendel approach [Hagras 2004], [Mendel 2001], [Wang 2003] to construct the initial rule base of the fuzzy system, which will be further optimized by the proposed BB-BC algorithm presented in the next subsection. The type-2 fuzzy system considered in this chapter extracts various multiple-input–multiple-output rules, which are known to model the relation between $M = (m_1, \ldots, m_p)$ and $O = (o_1, \ldots, o_q)$, and which use the following form:

$$IF\ m_1 \text{ is } \widetilde{X}_1^r \ldots \text{and } m_p \text{ is } \widetilde{X}_p^r \text{ THEN } o_1\ is\, \widetilde{Y}_1^r \ldots and\ o_q\ is\, \widetilde{Y}_q^r \quad (5.16)$$

where p is the amount of antecedents, q is the amount of consequents, r = *1*,…, R, R is the amount of the rules and r is the index of the current rule. We also have T_{in} interval type-2 fuzzy sets $\widetilde{X}_u^s$, $s = 1, \ldots, T_{in}$ for each input m_s where $u = 1,2, \ldots., p$ and T_{out} interval type-2 fuzzy sets $\widetilde{Y}_v^t$, $t = 1, \ldots, T_{out}$, for each output o_v where $v = 1,2, \ldots., q$.

For each training vector $(m^{(n)}; o^{(n)})$, $n = 1, \ldots, N$, where N is the amount the training date vectors while the upper membership degree and lower membership degree are calculated $\overline{\mu}_{\widetilde{X}_u^s}\left(m_u^{(n)}\right)$ and $\underline{\mu}_{\widetilde{X}_u^s}\left(m_u^{(n)}\right)$ for each fuzzy set of each input variable $\widetilde{X}_u^s$, $s = 1, \ldots, T_{in}$, $u = 1, \ldots, p$. After that, for each $s = 1, \ldots, T_{in}$, find $s^* \in \{1, \ldots, T_{in}\}$ such that [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$\mu^c_{\widetilde{X}_u^{s^*}}\left(m_u^{(n)}\right) \geq \mu^c_{\widetilde{X}_u^s}\left(m_u^{(n)}\right) \tag{5.17}$$

Where $\mu^c_{\widetilde{X}_u^s}\left(m_u^{(n)}\right)$ is the centre of the interval membership of $\widetilde{X}_u^s$ at $m_u^{(n)}$

$$\mu^c_{\widetilde{X}_u^s}\left(m_u^{(n)}\right) = \frac{1}{2}\left[\overline{\mu}_{\widetilde{X}_u^s}\left(m_u^{(n)}\right) + \underline{\mu}_{\widetilde{X}_u^s}\left(m_u^{(n)}\right)\right] \tag{5.18}$$

The following rule will be referred to as the rule generated by $(m^{(n)}; o^{(n)})$ [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$IF\ m_1\ is\ \widetilde{X}_1^{s^*(n)}\ and\ m_p\ is\ \widetilde{X}_p^{s^*(n)}\ THEN\ o\ is\ centered\ at\ o^{(n)} \tag{5.19}$$

An initial rule base will be constructed in this phrase. After that, conflicting rules which have the same antecedents but different consequents will be resolved by using the rule weight obtained by the following equation [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$w^{(n)} = \prod_{u=1}^{p} \mu^c_{\widetilde{X}_u^s}\left(m_u^{(n)}\right) \tag{5.20}$$

We will then divide the N rules into groups such that the rules in one group have the same antecedents so that [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$IF\ m_1 \text{ is } \widetilde{X}_1^r \text{ ... and } m_p \text{ is } \widetilde{X}_p^r \text{ THEN } o\ is\ centered\ at\ o^{(d_k^r)}$$

Where $k = 1, \dots, N$ and d_k^r is the data points index of group r. Then, the weighted average of the rules in group r whose amount of rule is N_r can be computed by using the following equation [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$\overline{w}^{(r)} = \frac{\sum_{k=1}^{N_r} o^{(d_k^r)} w^{(d_k^r)}}{\sum_{k=1}^{N_r} w^{(d_k^r)}} \tag{5.21}$$

After that, the conflicting rules in this group can be merged into one rule in the following format:

$$IF\ m_1 \text{ is } \widetilde{X}_1^r \text{ ... and } m_p \text{ is } \widetilde{X}_p^r\ THEN\ o\ is\ \widetilde{Y^r} \tag{5.22}$$

Where the choice of the output fuzzy set $\widetilde{Y^r}$ based is based on the following: among the T_{out} output fuzzy sets $\widetilde{Y^1}, \dots, \widetilde{Y^{T_{out}}}$ find the Y^{t*} such that [Hagras 2004], [Mendel 2001], [Wang 2003]:

$$\mu^c_{\widetilde{Y^{t*}}}\left(\overline{w}^{(r)}\right) \geq \mu^c_{\widetilde{Y^t}}\left(\overline{w}^{(r)}\right) \tag{5.23}$$

To expand the algorithm so as to handle multiple outputs, the steps of Equations (5.21), (5.22) and (5.23) are repeated for each output. An illustrative sample of fuzzy rules from our rule base could be seen in Table 5.3.

m_1	m_2	m_3	m_4	m_5	m_6	m_7	Outputs
LOW	MEDIUM	HIGH	MEDIUM	MEDIUM	LOW	MEDIUM	o_6 is High
LOW	LOW	MEDIUM	HIGH	LOW	HIGH	MEDIUM	o_4 is High
LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	o_1, o_3 is High
LOW	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	LOW	o_2 is High
MEDIUM	LOW	MEDIUM	HIGH	MEDIUM	HIGH	HIGH	o_5 is High
HIGH	LOW	LOW	MEDIUM	HIGH	MEDIUM	LOW	o_1, o_2 is High
LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	o_3 is High

Table 5.2: An illustrative sample of fuzzy rules of our rule base.

where the inputs are *left-arm-angle* (m_1), *right-arm-angle* (m_2), *left-hand-distance* (m_3), *right-hand-distance* (m_4), *body-bending-angle* (m_5), *spine-tofloor-distance* (m_6), *movement-speed* (m_7), and the outputs are *drinking/eating-possibility* (o_1), *sitting-possibility* (o_2), *standing-possibility* (o_3), *walking-possibility* (o_4), *running-possibility* (o_5), *lying/falling down–possibility* (o_6). For each rule in Table 5.3, in the outputs colums, the unshown outputs would have had an associated *LOW* fuzzy set.

5.4.2.7 Optimizing the IT2FLS via BB-BC

The main purpose of using FCM in order to generate the membership functions and use the Wang-Mendel method to construct the initial rule base before our BB-BC optimization, is to obtain a good starting point in the search space, since the BB-BC quality of the optimization highly relies on the starting state to converge fast to the optimal position. If we had started from random fuzzy sets and rules, the BB-BC would have taken a very long time to converge to optimal values.

The BB-BC optimization is an evolutionary approach which was presented by Erol and Eksin [Erol 2006]. It is derived from one of the theories of the evolution of the 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 from two phases: a Big Bang phase, where candidate solutions are randomly distributed over the search space in a uniform manner [Kumbasar 2011], and a Big Crunch phase, where candidate solutions are drawn into a single representative point via a centre of mass or minimal cost approach [Erol 2006]. All subsequent Big Bang phases are randomly distributed around the centre of mass or around the best fit individual in a similar fashion. The procedures followed in the BB-BC are as follows [Kumbasar 2011]:

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

$$x_c = \frac{\sum_{i=1}^{N} \frac{x_i}{f^i}}{\sum_{i=1}^{N} \frac{1}{f^i}} \tag{5.24}$$

where x_c is the position of the centre 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:

$$x^{new} = x_c + \frac{\gamma\rho(x_{max} - x_{min})}{k} \tag{5.25}$$

where γ is a random number, ρ is a parameter limiting search space, x_{min} and x_{max} are the lower and upper limits, and k is the iteration step.

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

5.4.2.7.1 Optimizing the rule base of the IT2FLS with BB-BC

To optimize the rule base of the IT2FLS, the parameters of the rule base are encoded into a form of a population. The IT2FLS rule base can be represented as shown below in Figure 5.7.

m_1^1	$\cdots$	m_p^1	o_1^1	$\cdots$	o_q^1	$\cdots$	m_1^R	$\cdots$	m_p^R	o_1^R	$\cdots$	o_q^R

Figure 5.45: The population representation for the parameters of the rule base

As showed in Figure 5.7, m_j^r are the antecedents and o_k^r is the consequents of each rule, respectively, where $j = 1,\ldots, p$, p is the number of antecedents; $k = 1,\ldots, q$, q is the number of the kinds of behaviour; $r = 1,\ldots, R$, and R is the number of the rules to be tuned. However, the values describing the rule base are discrete integers while the original BB-BC supports continuous values. Thus, instead of Equation (5.25), the following equation is used in the BB-BC paradigm to round off the continuous values to the nearest discrete integer values modelling the indexes of the fuzzy set of the antecedents or consequents.

$$D^{new} = D_c + round\left[\frac{\gamma\rho(D_{max} - D_{min})}{k}\right] \tag{5.26}$$

where D_c is the fittest individual, r is a random number, ρ is a parameter limiting search space, D_{min} and D_{max} are the lower and upper bounds, and k is the iteration step.

In this study, the rule base constructed by the Wang-Mendel approach [Mendel 2001], [Hagras 2007], [Wang 2003] 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 (5.25).

5.4.2.7.2 Optimizing the Type-2 membership functions with BB-BC

In order to apply BB-BC, the feature parameters of the type-2 membership function 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 so as to obtain σ_{k2}^{l} while σ_{k1}^{l} is provided by FCM. In order to be more accurate, the uncertainty factors α_k^j for each fuzzy set of the MFs are computed, where k = 1,…, p, p is the number of antecedents; j = 1,…, q, q is the number of input features. For illustration purposes, as in the MFs of the proposed system, three type-2 fuzzy sets including *LOW*, *MEDIUM* and *HIGH* are utilized for modelling each of the *7* features. Therefore, the total number of the parameters for the input type-2 MFs is *3*×*7*=*21*. In a similar manner, parameters for 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 Figure 5.8.

α_1^1	α_2^1	α_3^1	$\cdots$	α_1^7	α_2^7	α_3^7	α_L^{Out}	α_H^{Out}

Figure 5.46: 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 shown in Figure 5.8 and the constructed rule base shown in Figure 5.7, the recognition error in our solutions space can be minimized by using the following function as the cost function:

$$f^i = (1 - Accuracy^i) \tag{5.27}$$

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. The new candidates are generated using Equation (5.27).

5.5 Recognition Phase

In our fuzzy system, the antecedents are m_1, m_2, m_3, m_4, m_5, m_6, m_7 and each of these antecedents is modelled by three fuzzy sets: *LOW*, *MEDIUM*, and *HIGH*. The output of the fuzzy system is the behaviour possibility which is modelled by two fuzzy sets: *LOW* and *HIGH*. The type-1 fuzzy sets shown in Figure 5.5 have been obtained via FCM and the rules are the same as the IT2FLS.

When the system operates in real time, we measure $\{m_1, m_2, ..., m_7\}$ on the current frame and the IT2FLC is supposed to provide the possibilities of the candidate behaviour classes: *drinking/eating*, *sitting*, *standing*, *walking*, *running*, and *lying/falling down*. In our system, each activity category utilizes the same output membership function as depicted in Figure 5.11h, and product *t*-norm is employed while the centre of sets type-reduction for IT2FLS is used (for the compared type-1 FLS the centre of sets defuzzification is used). To recognize the current behaviour, our system works in the following pattern:

- The Kinect v2 is continuously capturing the raw 3D skeleton data from the subjects in the real-world intelligent environment,
- Then the raw real-time 3D Kinect data will be analysed by our feature extraction module to get the feature vector M = (m_1, m_2, m_3, m_4, m_5, m_6, m_7) modelling the behaviour characteristics in the current frame.
- For the crisp input vector M, a type-2 singleton fuzzifier will be used to fuzzify the crisp input and obtain the upper $\overline{\mu}_{\tilde{F}_1^i}(x')$ and lower ($\underline{\mu}_{\tilde{F}_1^i}(x')$) membership values.
- After that, we compute the firing strength $\underline{f}^i$ and $\overline{f}^i$ of each rule, where

$i = 1, \dots, R$, and R is the number of rules. Where $\underline{f}^i(x') = \underline{\mu}_{\tilde{F}_1^i}(x'_1) * \dots * \underline{\mu}_{\tilde{F}_p^i}(x'_p)$ and $\overline{f}^i(x') = \overline{\mu}_{\tilde{F}_1^i}(x_1) * \dots * \overline{\mu}_{\tilde{F}_1^i}(x'_p)$.

- The type reduction is carried out by using the KM approach to compute the type reduced set defined by the interval $[y_{lk}, y_{rk}]$.
- In the end, defuzzification is computed as $\frac{y_{lk}+y_{rk}}{2}$ to calculate the output degree of the target behaviour class. For one input feature vector analysed by our fuzzy system, we will have one output degree per candidate activity class, which models the possibility of the candidate activity class occurring in the current frame.

In our application scenario within AAL spaces, we aim at recognizing the regular daily activities. However, the subject's activity sequence happening at the actual environment is not a continuous time-series due to the occlusion problems, capturing angle, and the casualness of the subject which could lead to untargeted and unknown types of behaviour which are out of our concern range. To solve this problem, we do not use shoulder functions in our membership functions since the target behaviours are only modelled by the feature values ranging in the sections returned by FCM learned from the feature data of the concerned activities. Additionally, we will check if the behaviour candidate is confident in the current frame by checking if its associated output degree is higher than a confidence threshold *t*; in our experiment we set *t* = 0.62. The confident behaviour candidates will be further considered to get the final recognition output.

In our application scenario, some of the target behaviour categories are conflicting because it is impossible for them to be happening at the same moment.

Therefore, in our experiment, we divide the target behaviour categories into several conflicting groups, *sitting*, *standing*, *walking*, *running*, and *lying*/*falling down* are part of the same group while *drinking/eating* is another group.

In the final step, the behaviour recognition is performed by choosing the confident candidate behaviour category with the highest output degree as the recognized behaviour class in its behaviour group. For example, if the outputs of *sitting*, *standing*, *walking*, *running*, and *lying*/*falling down* are *0.25*, *0.75*, *0.64*, *0.0*, *0.0*, respectively, and the output of *drinking/eating* is *0.25*, then the final recognition result would be *standing* since its output degree is the highest amongst the confident candidates (which are *standing and walking* in this case) in its group and the output degree of *drinking/eating* in the other group is lower than the confident level. However, in a very rare situation that two confident candidate categories in a conflicting group are allocated with a same output degree, this demonstrates that the two candidates have extremely high behavioural similarity and cannot be distinguished in the current frame. In that case our system ignores these two candidate categories in the behaviour recognition of the current frame.

5.6 The GUI of the Proposed System

The system detects six kinds of behaviour which are crucial for AAL activities. These are falling down, drinking/eating, walking, running, sitting and standing.

The GUI of the proposed system has two parts. The first part is shown in Figure 5.9. It is used during the video capture and shows the detected behaviour and it can send immediate alerts for important incidents such as falling down. The left part of Figure 5.9 shows the original colour high-definition video which is continuously captured and displayed. The right part of Figure 5.9 shows the captured 3D skeleton data (highlighted in green) of the subject in the current frame. The GUI shows also the detected behaviour for up to six users in the current frame. As can be shown in Figure 5.9, the system was able to detect the event of “falling down/lying down” (highlighted with the red rectangle) under strong sunshine illumination and with the existence of shadow changes.

The second part of the GUI is shown in Figure 5.10 and it deals with the event retrieval, linguistic summarisation and playback. In the initial procedure of the GUI, the connection between the GUI to the back-end event SQL server will be built automatically. After that, the user can search for the events that are of interest by entering their search criteria including the identification of the subject, the number of the subject, event category, and event timestamp. An example has been given in Figure 5.10 where the user has selected to search the event category “Fallingdown” from the target behaviour list as highlighted with the red rectangle in Figure 5.10a. For further refinement of the retrieval criteria, the particular subject number as well as a fixed time period described by the exact starting date and time and the ending date and time of the event timestamp has been also provided by the user, as highlighted with the red rectangle in “Event Retrieval” section of Figure 5.10b. After clicking the “Retrieve” button, the front-end GUI will translate the current searching criteria into SQL scripts as shown in the edit box “SQL script” (for further editing of complex and advanced searching if necessary). Then the translated SQL scripts will be sent from the front-end GUI to the back-end event database server to retrieve the relevant events according to the requests of the user. Finally, the retrieved events with details including subject information, event descriptions, and the relevant video clips will be sent from the back-end event server to the front-end GUI. The results of event retrieval are depicted in the list showing the relevant activities which have previously been detected and stored, as shown in Figure 5.10b. The details of the selected event in the retrieval list is shown in the event information section, and the retrieved events can be used to playback the video matching the sequences the user wants to see as shown in Figure 5.10c.

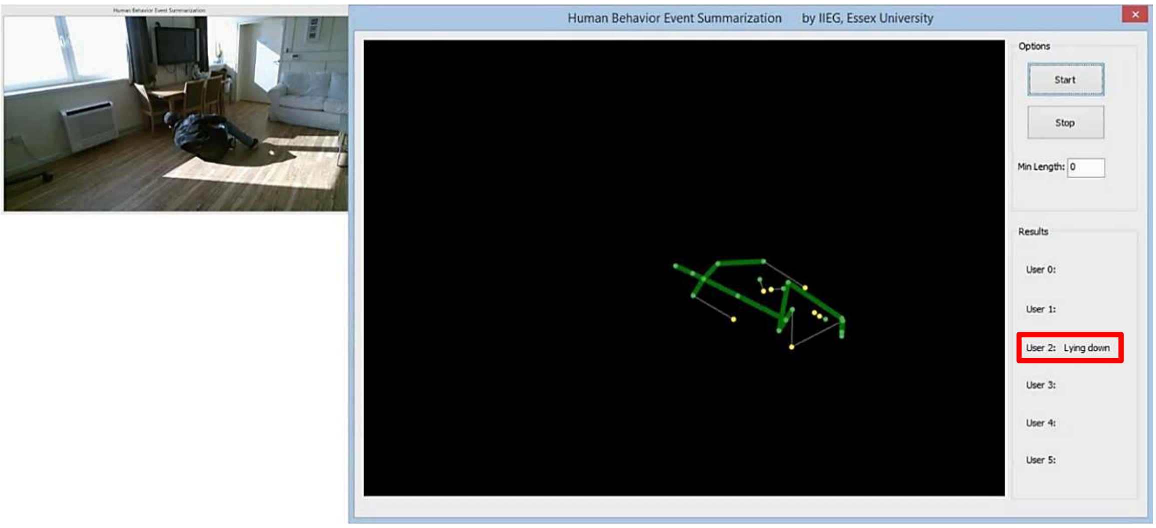


Figure 5.47: The event detection GUI.

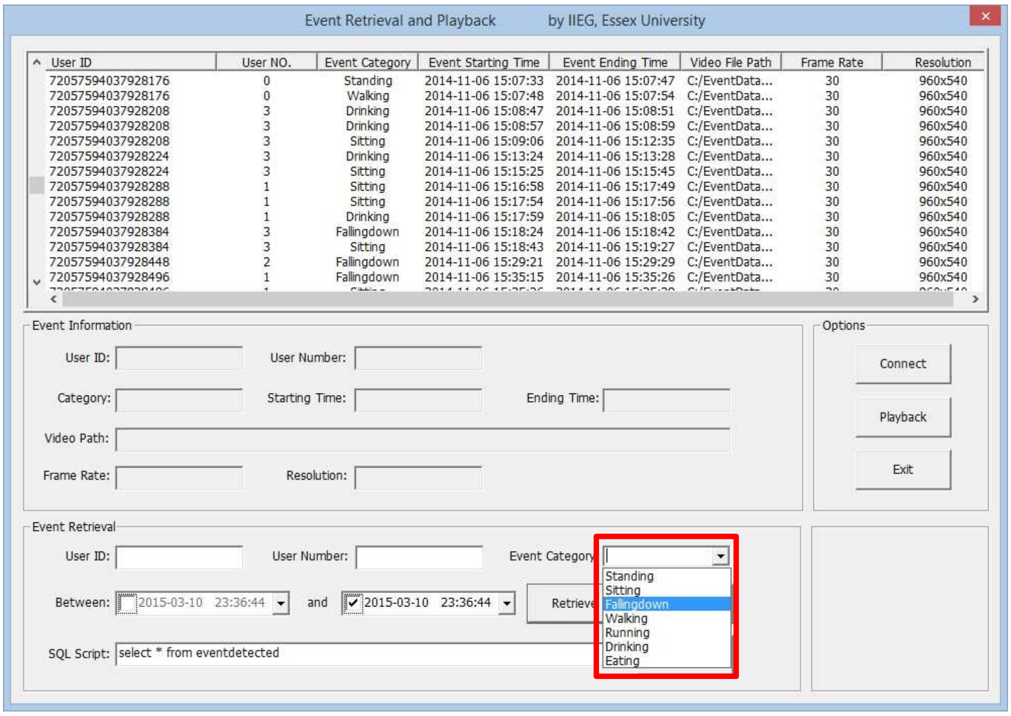
5.7 Event Summarisation

Based on the results of the behaviour recognition, the event summarisation will be performed by combining all the related key information and summarise them into an event which will then be sent and stored into the back-end event database. Since this event detection system is connected to the back-end event database, once an activity is detected, the system will summarize the relevant details of an event (e.g. subject identification, subject number, behaviour category, event time stamp, event video data, etc.) regarding the detected behaviour. After that, the summarized event will be sent to the back-end SQL event database and will be efficiently stored so that event retrieval and playback can later be performed by the users using the front-end

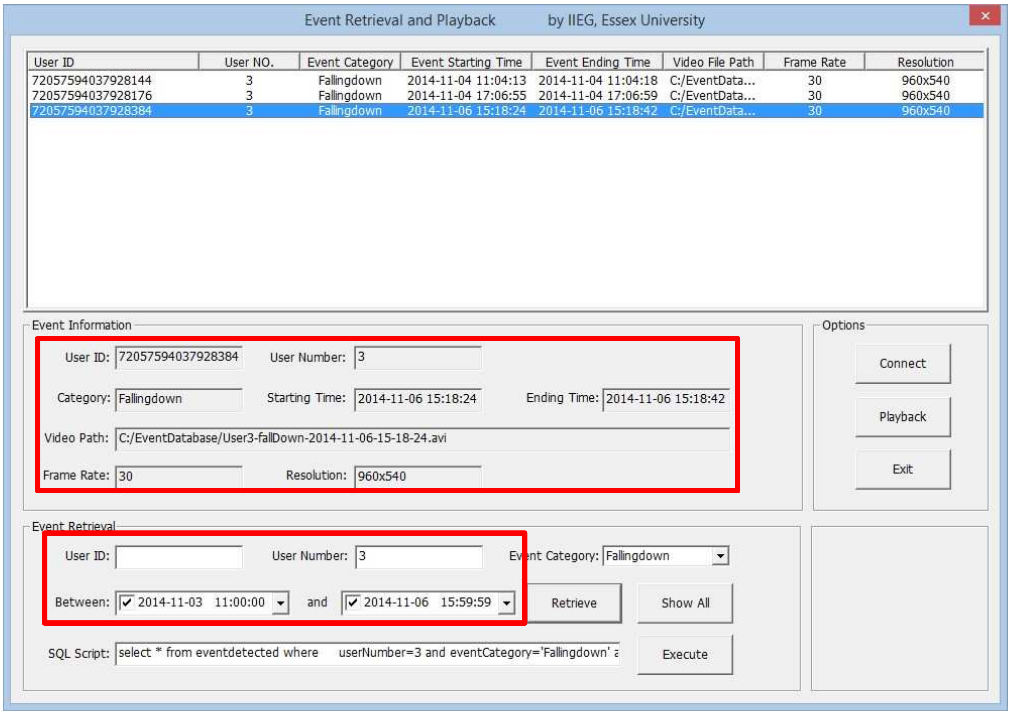
GUI system. Meanwhile, if the detected event is an emergency, a warning message will be sent to relevant caregivers so that instant action can be taken.

5.8 The Back-end Server Event SQL Database of the Software System

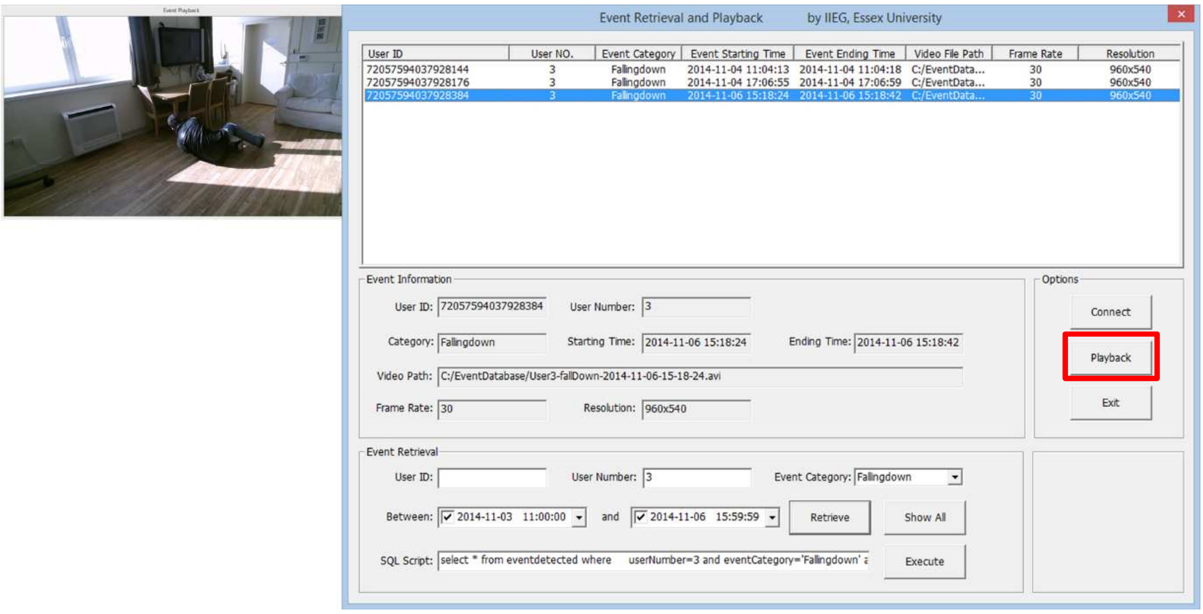
The back-end event database is developed for the efficient storage of the detected events including event details such as subject identification, subject number, event category, event starting time, event ending time, and the associated high-definition video of the event. The event SQL database provides the services of event search and retrieval for different front-end user interfaces to locally or remotely retrieve and play back the interesting events.



(a)



(b)



(c)

Figure 5.48: The front-end GUI for the event search, lingustic summarisation and video playback (a) Candidate event categories for event retrieval (b) Other search criteria for event retrieval and the details of the selected event from the retrieval list (c) Event playback

5.9 Experiments and Results

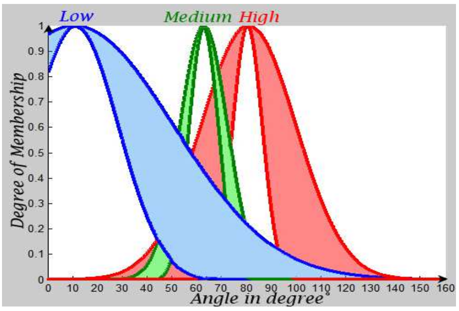
In our application scenarios, we aim at recognizing the following types of behaviour: *drinking/eating, sitting, standing, walking, running, and lying/falling down*. Our experiments were performed in different places such as the intelligent apartment (iSpace) [Hagras 2007] and intelligent Classroom iClassroom [Almohammadi 2014] at the University of Essex. The purpose of choosing these behaviours as target categories is that they are common behaviours and activities in AAL environments. We detect fall-down event so that our system can send a warning message to the nearby caregivers or other people who can help [Popoola 2012] [Nait-Charif 2004]. And our system summarises the frequency of the drinking activity to ensure that the user drinks enough water throughout the day to avoid dehydration [Tham 2014] [Maierdan 2013]. By a daily summarization of the sitting and lying duration and frequency, healthcare advice would be provided if the user remains inactive/active most of the time [Barnes 1998] [Nambu 2008]. The detection results of running demonstrate a potential emergency happening [Foroughi 2008]. From the detection results of standing and walking, our system obtains the location and trajectory of the subject so that services, such as wandering prevention, can be provided to dementia patients [Wan 2011]. Also, the risk of falling down can be reduced by analysing the pattern of standing and walking [Wu 2008]. Furthermore, cognitive rehabilitation services can be provided to help the elderly with dementia by summarizing this series of daily activities [Hoey 2010] [Levinson 1997].

We tested our methods using the Type-1 Fuzzy Logic System (T1FLS) and Type-2 Fuzzy Logic System (T2FLS) compared against the non-fuzzy traditional methods including Hidden Markov Models (HMM) and Dynamic Time Warping (DTW) on 15 subjects ensuring high-levels of intra- and inter- subject variation and ambiguity in behavioural characteristics. In this 3D experiment, we don't use Weizmann dataset anymore, as the Weizmann human action dataset [Blank 2005] was firstly published in 2005 and finally revised in 2007, and the main purpose of the Weizmann dataset is to fairly benchmark algorithms for behaviour recognition on CCTV camera or video sequence, so the data in Weizmann dataset are all two dimensional. Another reason for that fact that Weizmann dataset doesn't contain three dimensional data is that there was no 3D video sensor in 2005 and the first 3D video sensor (Kinect v1) with development kit was released in July 2011.

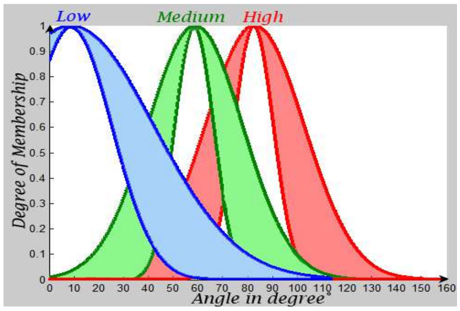
In the training stage, the training data were produced by different subjects. The subjects were asked to perform each target behaviour on average two to three times. This resulted in around *220* activity samples for training. In the real-world recognition stage, we divided the subjects into different groups and we performed the experiments with different subject numbers in a scene to model different uncertainty complexity. The experiments were conducted on average with five repetitions per target behaviour by each subject in the group analysed by the real-time behaviour recognition system. This resulted in around *1,600* activity samples for testing. To perform a fair comparison, all methods share the same input features. As in real-world environments, occlusion problems exist in our test cases, which lead to behavioural uncertainty caused by the occlusions of the subjects. The experiments were conducted with different subjects and different scenes in various circumstances including different illumination strength, partial occlusions, daytime and night time, moving camera, fixed camera and different monitoring angles, among others. The experiment results demonstrate that our algorithm is robust and effective in handling the high levels of uncertainties associated with real-world environments including occlusion problems, behaviour uncertainty, activity ambiguity, and uncertain factors such as position, orientation and speed and so on.

The type-2 membership functions used in our system, which are constructed and optimized by BB-BC, are shown in Figure 5.11.

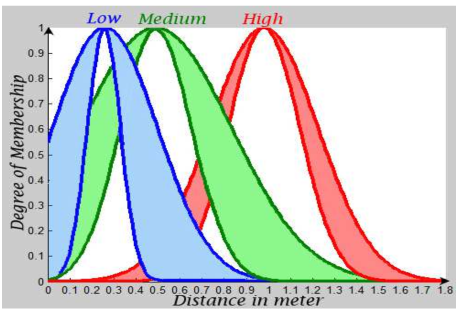
Our experiment result demonstrates that the BB-BC optimization improves the performance of our type-2 fuzzy logic system. In the BB-BC optimization procedure of the type-2 membership functions, we set x_{min} and x_{max} to 50% and 300%, which influences the FOU blurring factor α in type-2 MFs construction. In order to achieve robust recognition performance, in our experiment the population size N of BB-BC is set to 200,000. In addition, owing to the high-performance of BB-BC, each iteration of the optimization procedure can be conducted in a few minutes.



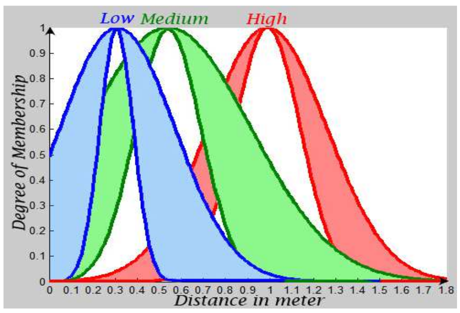
(a)



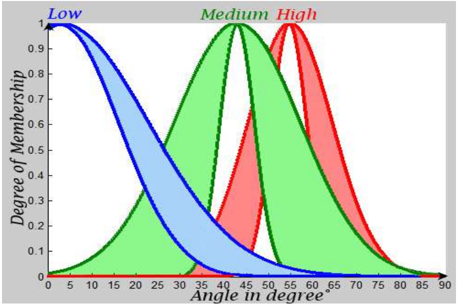
(b)

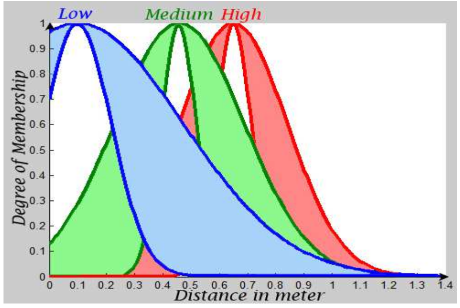


(c)

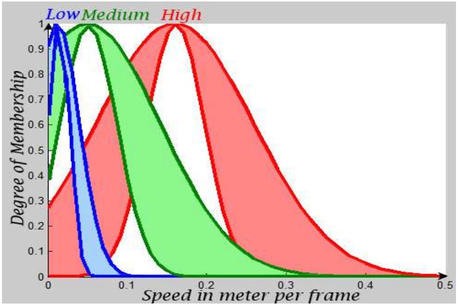


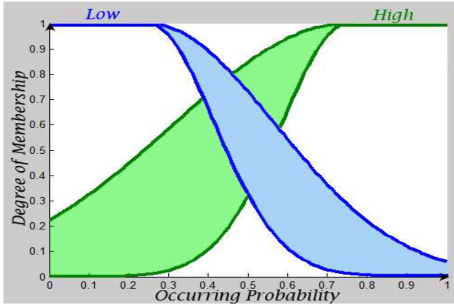
(d)





(e) (f)





(g) (h)

Figure 5.49: (a) Type-2 MF for m_1 (b) Type-2 MF for m_2 (c) Type-2 MF for m_3 (d) Type-2 MF for m_4 (e) Type-2 MF for m_5 (f) Type-2 MF for m_6 (g) Type-2 MF for m_7 (h) Type-2 MF for Output

Based on the optimized type-2 fuzzy sets and rule base by utilizing BB-BC, our IT2FLSs-based system outperforms its counterpart T1FLSs-based recognition system (see Table 5.4), where the type-2 system achieves *5.29*% higher average per-frame accuracy than the type-1 system over the test data in the recognition phrase. Our type-2 fuzzy logic system also outperforms the traditional non-fuzzy based recognition methods based on Hidden Markov Models (HMM) [Kim 2010] and Dynamic Time Warping (DTW) [Reyes 2011]. In order to make a fair comparison with the traditional HMM-based and DTW-based methods, all methods share the same input features. As shown in Table 5.4, our IT2FLSs-based method with a BB-BC optimization achieves *15.65*% higher recognition average accuracy than the HMM-based algorithm, and it also achieves *11.62*% higher recognition average accuracy than the DTW-based algorithm. For the standard deviation of each subject's recognition accuracy, the T2FLS-based method is the lowest, demonstrating the stableness and robustness of our method when testing different subjects.

When the number of subjects increases, this leads to a higher possibility of occlusion problems with a higher-level of behaviour uncertainty, and the difference between our method compared to the T1FLS-based method and the traditional non-fuzzy methods is even higher, as shown in Table 5.5, Table 5.6 and Table 5.7. Our T2FLS-based method remains the most robust algorithm with the highest recognition accuracy which remains roughly the same when adding more users to the scene.

Based on the recognition results of our optimized IT2FLS, higher-level applications including video linguistic summarization, event searching, activity retrieval, event playback, and human-machine interactions have been developed and successfully deployed in iSpace and iClassroom. In our experiments of this chapter, fifteen human subjects were involved in one thousand eight hundred twenty activity sequences. And we performed fifteen sessions run to analyse to the results. In our experiments of this chapter, free behaviour was used in all of the tests.

Method	Average Accuracy	Standard Deviation
HMM	70.9266%	0.175258
DTW	74.9614%	0.129266
T1FLS	81.2903%	0.110410
T2FLS	86.5798%	0.086551

Table 5.3: Comparison of Fuzzy-based methods against traditional methods with One subject per Group in a scene (Fifteen groups)

Method	Average Accuracy	Standard Deviation
HMM	72.4134%	0.078800
DTW	71.6549%	0.051693
T1FLS	79.0394%	0.157738
T2FLS	85.8864%	0.092471

Table 5.4: Comparison of Fuzzy-based methods against traditional methods with Two subjects per Group in a scene (Six groups)

Method	Average Accuracy	Standard Deviation
HMM	70.1782%	0.042738
DTW	73.7452%	0.103744
T1FLC	78.3855%	0.128380
T2FLC	86.1305%	0.082625

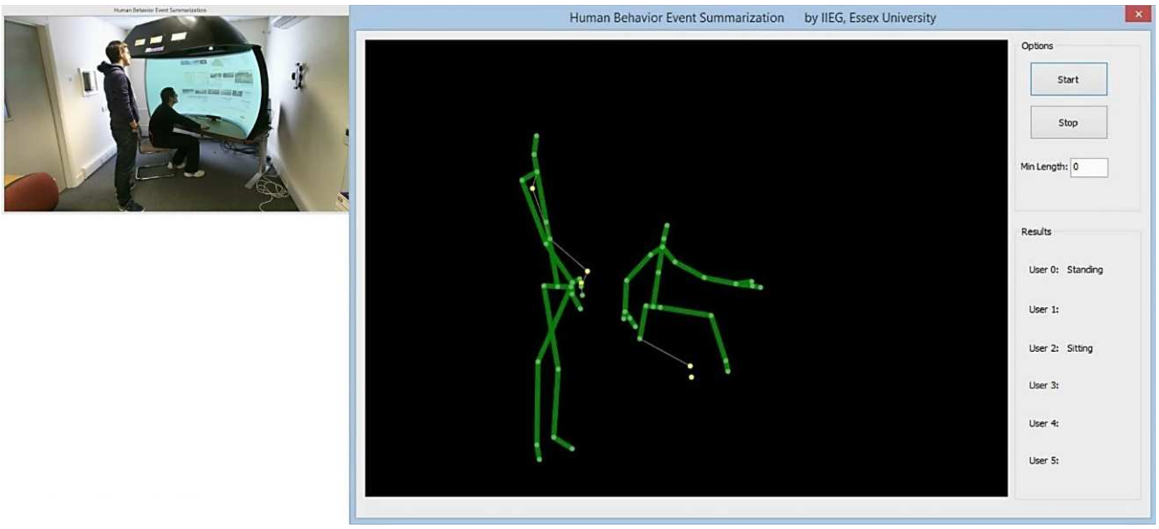
Table 5.5: Comparison of Fuzzy-based methods against traditional methods with Three subjects per Group in a scene (Five groups)

Method	Average Accuracy	Standard Deviation
HMM	69.5274%	0.083920
DTW	70.1220%	0.112780
T1FLC	76.6017%	0.080618
T2FLC	84.7253%	0.072113

Table 5.6: Comparison of Fuzzy-based methods against traditional methods with Four subjects per Group in a scene (Three groups)

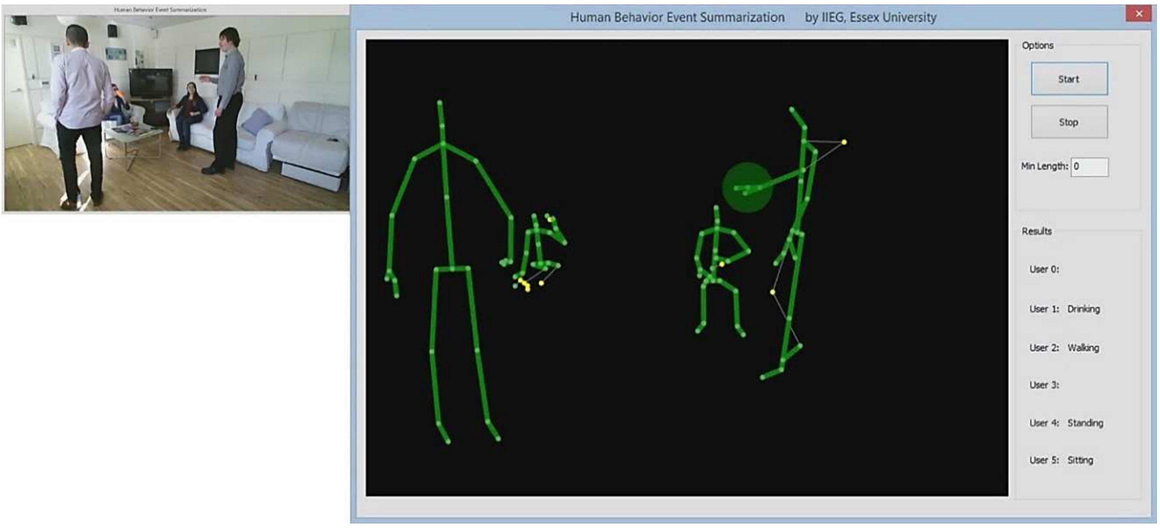
The results of detected events and the associated video data are stored in the SQL Event database server so that further data mining can be performed by using our event summarization and retrieval software. Furthermore, the user can easily summarize the event of interest at the given time frame and play them back.

Figure 5.12 provides the detection results of the real-time event detection system deployed in different real-world intelligent environments such as iClassroom and iSpace at the University of Essex. The number of subjects changes according to the application scenario. In Figure 5.12a, two students are using our immersive learning platform [Rios 2013] in iClassroom with one Kinect v2. In Figure 5.12b, the system analysed the activity of the three subjects in the scene in the iClassroom. In Figure 5.12c, behaviour recognition is performed in the iSpace with four subjects. As the scenario is in a living environment, the users have more freedom to act casually and occlusion problems are more likely to occur with a large crowd of subjects. Consequently, there are higher-levels of uncertainty. As it can be seen, user 1 who is drinking coffee is heavily occluded by the table in front, and so is user 2 who is walking towards the door. Our IT2FLS-based recognition system handles the high-levels of uncertainty robustly and returns the correct results.



(a)

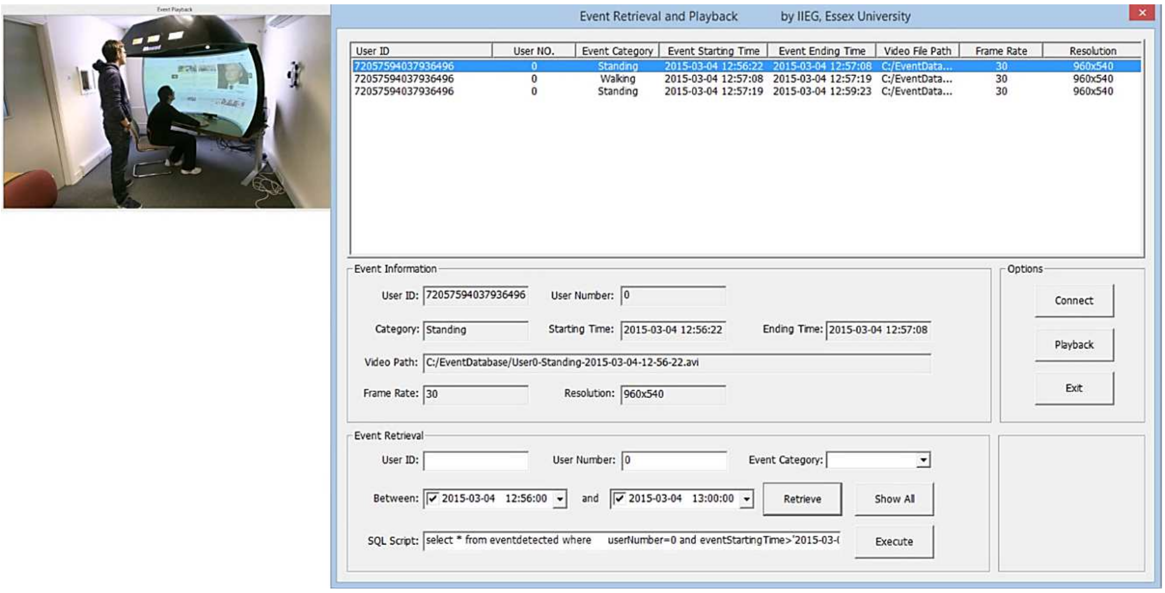
(b)



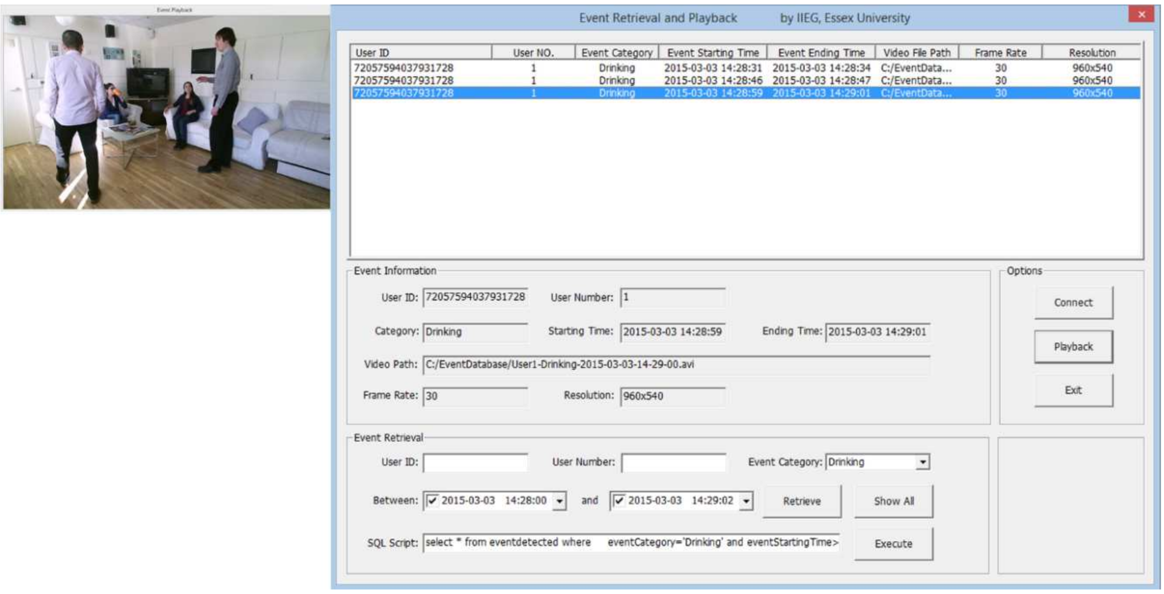
(c)

Figure 5.50: Detection results from our real-time IT2FLS-based recognition system, (a) recognition results in the iClassroom with two subjects in the scene (b) recognition results in the iClassroom with three subjects in the scene (c) recognition results in the iSpace with four subjects in the scene leading to occlusion problems and high-levels of uncertainty

As shown in Figure 5.13a, in order to retrieve any interesting events conducted by a certain subject during a fixed time period, we inputted a subject number and time duration, and performed event retrieval via the front-end GUI. After that, the relevant retrieved events were shown in the result list, from which we selected the retrieved event and played back the HD video. Similarly, in Figure 5.13b, we were interested in the drinking activities that happened in the iSpace. Therefore, we selected the "Drinking" activity from the event category and also provided a certain time period. Then, the events associated with "Drinking" in the given time period were retrieved.



(a)



(b)

Figure 5.51: Event retrieval and playback, (a) Event retrieval and playback with a given subject number and time period in the iClassroom (b) Event retrieval and playback with a given event category and time period in the iSpace

5.10 Discussion

In this chapter, we presented a robust behaviour recognition algorithm for video linguistic summarization using a 3D Kinect camera based on Interval Type-2 Fuzzy Logic Systems. 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 algorithm. Our experiments have been successfully conducted in real-world intelligent environments. The experiment results show that the proposed IT2FLS outperformed its 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.

Chapter 6: Conclusions and Future Work

This thesis starts with the introduction in Chapter 1 which presented the background of event detection and summarization and the corresponding applications in intelligent environments and ambient assisted living. Event detection and summarization is used to detect important and interesting information from massive raw data captured by the sensors. In this way further data mining can be done to summarise the detection results into events from which the user can easily and quickly browse core information. After that, we explained the problem of high-levels of uncertainties existing in ambient assisted living caused by the noise factors associated with the real-world environments. Then we presented the application of fuzzy sets in event detection and summarization which can handle the uncertainty and achieve robust detection. Next we introduced the objectives, the novelty and significance, and the structure of this thesis.

In Chapter 2, we laid out the theoretical background of this work. We firstly introduced the basic concepts and theory of a fuzzy logic system and the relevant concepts including fuzzy logic, fuzzy set and its operations and linguistic variables. Subsequently, we explained the type-1 fuzzy logic system and interval type-2 fuzzy logic system as well as their main calculation procedures. We also explained that type-1 fuzzy logic are not robust enough for handling the high-level of uncertainties associated with the real-world environment. Therefore we concluded that there is a demand for a system that is capable of robustly handling these uncertainties automatically and adaptively in order to adjust to the variations and changes in real-world environments. To perform automatic optimization, we reviewed the recent popular methods for optimizing fuzzy logic system. We introduced the basis of big bang-big crunch which is capable of tuning the parameters of the fuzzy membership functions and the rule base of the fuzzy logic system. Finally the fuzzy logic system will use these optimized parameters so that the optimum performance and accuracy can be achieved. Then, we discussed the conventional works in the field of 2D human behaviour recognition and detection.

In Chapter 3, we presented a type-2 fuzzy logic based system for robustly extracting the human silhouette which is a fundamental and important procedure for advanced video processing applications, such as pedestrian tracking, human activity analysis, and event detection. Due to the huge complexity of a dynamic and real-life environment, the problem of detaching moving objects from human silhouette is complicated. In order to address this problem without high computational complexity, we firstly use GMM to detect the foreground and then we employ an IT2FLS for objects detachment. We conducted real-world experiments which have shown that the proposed IT2FLS is effective in detaching objects; any misclassifications are greatly reduced compared to the similar T1FLS. At the same time, the IT2FLS results also high in accuracy for silhouette extraction compared to the T1FLS. Hence, by utilizing IT2FLS the proposed system achieves a silhouette extraction with good robustness against noise factors and uncertainties:

- Light condition changes.
- Reflection of human body.
- Moving objects attached to the human silhouette, etc.

In order to demonstrate the robustness of the proposed system, plenty of experiments have been performed in various environments such as indoor environments and outdoor environments in different situations:

- Single subject at the scene.

- Multiple subjects at the scene.
- Strong sunshine (illumination) at the scene.
- Snowy weather in the scene.

For comparison purposes, in our experiment results it can be seen clearly in our experiment results that the misclassification of the proposed IT2FLS is significantly reduced compared to T1FLS, while the IT2FLS also results in higher accuracy than the T1FLS in silhouette extraction. In the experiment of single subject indoors, T2FLS improves the accuracy by *6.74*% and reduce the average misclassification by *125.13* pixels comparing against T1FLS; When we move the test environment outdoors, the difference between T1FLS and T2FLS became larger, in the multiple subjects outdoors, T2FLS improves the accuracy by *4.76*% and reduce the average misclassification by *220.84* pixels comparing against T1FLS. And the gap was even enlarged when the environment is snowy causing more uncertainty. In the multiple subjects in outdoor snow environment, T2FLS improves the accuracy by *13.3*% and reduce the average misclassification by *387.46* pixels comparing against T1FLS. And in the last experiment, when the environment was crowded with people causing high-level of uncertainties, T2FLS improves the accuracy by *9.69*% and reduce the average misclassification by *624.78* pixels. Silhouette extraction the fundamental and core task in advanced smart vision system for event detection and summarization.

Based on the extracted silhouette by our proposed algorithm as the starting module in the smart vision system, in Chapter 4 we proposed a computationally efficient fuzzy logic based system for the automatic recognition of human behaviour using machine vision for applications in intelligent environment. It is hoped that the proposed method will be an enabling step towards the realization of ambient intelligent environments which can automatically detect human behaviour and adapt the user's environment accordingly. To the best of the author's knowledge, this thesis is the first in the literature applying fuzzy logic systems for visual-based humans' behaviour recognition. In so doing, the original images are first captured from the input video sequences and the extracted human silhouette is generated using our proposed method based on an interval type-2 fuzzy logic system. After that, the input features are computed from the extracted silhouette images using a seven-dimensional model-based feature set including motion information and shape descriptors. Finally, human behaviour is recognized based on the input feature set by using the proposed fuzzy-based recognition method.

We have successfully tested our system on the publicly available Weizmann human action dataset [31], where our fuzzy based system produced an robust recognition accuracy of *100*%, which outperformed the traditional non-fuzzy systems based on hidden Markov models by an enhancement of *13.3*% accuracy. It also outperformed the recognition accuracy of other state-of-the-art approaches including hCRF-based method and codebook-based algorithm, which were also applied on the Weizmann dataset, by *9.71*% and *45*%, respectively. Also it is important to note that the IT2FLSs-based system outperforms its T1FLSs-based recognition system counterpart with the same rule base (either manually designed or optimized by BB-BC). Our manually designed IT2FLSs-based method (using manual rule base and α=*18*%) achieves *4.09*% higher average per-frame accuracy than the manually designed T1FLS. In addition, the BB-BC optimized IT2FLS achieves *4.55*% higher average per-frame accuracy than the BB-BC optimized T1FLS. The per-video accuracy of the proposed method *100*% outperforms the traditional non-fuzzy approaches including hCRF-based method [5], SVM-based approach [6] and EM-based algorithm [3] by *2.78*%, *1.20*% and *27.20*%, respectively. Moreover, our system provides a relatively computationally efficient and robust response since our method can process *30* frames per second, which improves *100*% of the analysis speed when compared to the HMM-based algorithm in [25], which can only process *15* frames per second. We have also outperformed the computation speed of SVM-based approach in [37] by *6876.74*% whereas the SVM approach can only process *0.43* frames per second. The experiment results demonstrate the superior performance of the proposed BB-BC based approach for optimizing the membership functions and rule base of the IT2FLS which outperformed the equivalent T1FLS and the state-of-the-art non-fuzzy methods regarding recognition accuracy and analysis performance.

In Chapter 5 and in relation to ambient assisted living, we introduced a framework for behaviour recognition and event linguistic summarization utilizing a RGB-D sensor Kinect v2 based on BB-BC optimised Interval Type-2 Fuzzy Logic Systems (IT2FLSs) for AAL real world environments. We showed that the proposed system is capable of handling high-levels of uncertainties caused occlusions, behaviour ambiguity and environmental factors. In the proposed system, the input features were first extracted from the 3D Kinect data captured by the RGB-D sensor. After that, membership functions and rule base of the fuzzy system were constructed automatically based on the obtained feature vectors. Finally, we used a Big Bang–Big Crunch (BB-BC) based optimization algorithm to tune the parameters of the fuzzy logic system for behaviour recognition and event summarization. For the real-world application in AAL environments, we developed a real-time distributed analysis system including front-end user interface software for operational commands inputting, a real-time learning and recognition system to detect the users' and a back-end SQL database event server for smart event storage, high-efficient activity retrieval, and high-definition event video playback. Our proposed system has been successfully deployed in real world environments occupied by various users ensuring a high-level of intra- and inter- subject behavioural uncertainty. Our experiment results demonstrated that the BB-BC based optimization paradigm is effective in tuning and optimizing the parameters of our fuzzy system. In addition, our experiment results with single users show that the proposed IT2FLS handles the high-level of uncertainties well and achieves robust recognition of *86.57*% and it also outperformed its T1FLS counterpart by an enhancement of *5.28*% as well as other traditional non-fuzzy systems, including the HMM-based system and DTW-based method by *15.65*% and *11.61*%, respectively. Moreover, it was shown that the proposed IT2FLS delivers consistent and robust recognition accuracy while the T1FLS and other conventional methods based on HMM and DTW show degradations in recognition accuracy when one increases the number of users.

For our on-going research, we will focus on applying our system in different applications such as ambient assisted living and healthcare. We will explore other types of hardware platforms and sensors to build a low-cost but practical solution for a wider and scale-up application range. In the side of functionality, we will investigate in the application of person identification using the technologies of face recognition, and object identification to identify the users in crowded environment and re-identify the human subject when he re-enters the scenario. And we will develop this new module into our current system and allows further human-machine interaction for example to communicate with human user and improve the user experience and satisfaction. Also, we will develop more user interface system which allows the users to configure the parameters of the target behaviour/event. In the other side of algorithm, we will investigate in the state-of-the-art algorithms such as support vector machine, random forest classifier, deep neural network, etc. As the complexity of the problem of human identification, behaviour recognition and event summarisation, we will explore and benchmark the state-of-the-art techniques and software libraries in machine learning and deep learning (e.g., Torch, Theano, Caffe, and SciKitLearn) in our system and try to utilize the modern big-data computing ecosystems such as Apache Spark in mining the massive data from the real-world environment.

References

[Acampora 2015] G. Acampora, P. Foggia, A. Saggese, M. Vento (2015): A Hierarchical Neuro-Fuzzy Architecture for Human Behavior Analysis. In: Information Sciences, 2015.

[Acampora 2012] G. Acampora, P. Foggia, A. Saggese, M. Vento (2012): Combining Neural Networks and Fuzzy Systems for Human Behavior Understanding. In: Proceedings of the "IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)", pp. 88-93, IEEE, 2012, ISBN: 978-1-4673-2499-1

[Aubert 2001] G Aubert and P. Kornprobst, “Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations,” Applied Mathematical Sciences, vol. 147, Springer, New York, 2001.

[Afshar 2011] M. Afshar, and I. Motaei, “Constrained big bang-big crunch algorithm for optimal solution of large scale reservoir operation problem,” Int. J. Optim. Civil Eng., vol. 2, pp. 357-375, 2011.

[Aghajan 2007] H. Aghajan, R. Kleihorst (Eds.), in: Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras, 2007.

[Aghajan 2009] H. Aghajan, A. Cavallaro, “Multi-Camera Networks: Principles and Applications,” Academic Press, 2009.

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

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

[Anderson 2009] 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.

[Alcalá 2007a] R. Alcalá, J. Alcalá-Fdez, and F. Herrera, "A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection," IEEE Transactions on Fuzzy Systems, vol. 15, no. 4, pp. 616–635, 2007.

[Alcalá 2007b] R. Alcalá, M.J. Gacto, F. Herrera, and J. Alcalá-Fdez, "A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems," International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, vol. 15, no. 5, pp. 539–557, 2007.

[Ahad 2011] M. Ahad, J. Tan, H. Kim, and S. Ishikawa, "Action dataset - a survey," SICE Annual Conference, SICE, pp. 1650-1655, 2011.

[Almohammadi 2014] K. Almohammadi, B. Yao, and H. Hagras, "An interval type-2 fuzzy logic based system with user engagement feedback for customized knowledge delivery within intelligent E-learning platforms," Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 808-817, 2014.

[Almohammadi 2015] K. Almohammadi, B. Yao, and H. Hagras, "An Interval Type-2 Fuzzy Logic Based System for Improved Instruction within Intelligent E-Learning

Platforms," Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015.

[Aldridge 2009] B. Aldridge, J. Saez-Rodriguez, J. Muhlich, P. Sorger, and D. Lauffenburger "Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling," Public Library of Science Computational Biology, vol. 5, no. 4, 2009.

[Barnes 1998] N. Barnes, N. Edwards, D. Rose, and P. Garner, "Lifestyle monitoring technology for supported independence," Computing & Control Engineering Journal, vol. 9, pp. 169–174, Aug. 1998.

[Blackman 1999] S. Blackman, R. Popoli, "Design and Analysis of Modern Tracking Systems," Boston, MA: Artech House, 1999.

[Blank 2005] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," International Conference on Computer Vision, ICCV, pp. 1395-1402, 2005.

[Bobick 2001] A. Bobick and J. Davis, "The representation and recognition of action using temporal templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257–267, 2001.

[Black 1997] M. Black, Y. Yacoob, A. Jepson, and D. Fleet, "Learning parameterized models of image motion," IEEE International Conference on Computer Vision and Pattern Recognition, pp. 561–567, 1997.

[Bobick 2001] A. Bobick and J. Davis, “The representation and recognition of action using temporal templates,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257–267, 2001.

[Brand 1997] 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.

[Callaghan 2004] V. Callaghan, G. Clark, M. Colley, H. Hagras, JSY. Chin, and F Doctor. “Inhabited intelligent environments,” BT Technology Journal, vol. 22, no. 3, pp. 233–247, July 2004.

[Cirstea 2002] M. Cirstea, A. Dinu, M. McCormick and J.G. Khor, “Neural and Fuzzy Logic Control of Drives and Power Systems,” ISBN: 0-7506-5558-5, September 2002.

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

[Chen 2006] X. Chen, Z. He, J. Keller, D. Anderson, and M. Skubic, “Adaptive Silhouette Extraction in Dynamic Environments Using Fuzzy Logic,” Proceedings of the World Congress in Computational Intelligence, WCCI, 2006.

[Collins 2001] R. Collins, A. Lipton, H. Fujiyoshi, T. Kanade, “Algorithms for Cooperative Multi-sensor Surveillance,” in: Proceedings of the IEEE, vol. 89, pp. 1456–1477, 2001.

[Cho 2008] K. Cho, I. Hwang, S. Kang, B. Kim, J. Lee, S. Lee, S. Park, J. Song, Y. Rhee, “HiCon: A Hierarchical Context Monitoring and Composition Framework for Next-generation Context-aware Services,” IEEE Transaction on Network, vol. 22, no. 4, pp.34-42, 2008.

[Chi 1996] Z. Chi, H. Yan, T. Pham, “Fuzzy Algorithms: With Applications to Image Processing and Pattern recognition,” World Scientific, Singapore, 1996.

[Cordon 2004] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, “Ten years of genetic fuzzy systems: current framework and new trends,” Fuzzy Sets and Systems, vol. 141, no. 1, pp. 5–31, 2004.

[Casillas 2005] J. Casillas, O. Cordon, M.J. del Jesús, and F. Herrera, “Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction,” IEEE Transaction on Fuzzy Systems, vol. 13, no. 1, pp. 13–29, 2005.

[Dalal 2005] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Conference on Computer Vision and Pattern Recognition, CVPR, pp. 20-25, 2005.

[Davis 2006] J. Davis and A. Tyagi, “Minimal-latency human action recognition using reliable-inference,” Image and Vision Computing, vol. 24, no. 5, pp. 455–472, 2006.

[Dollar 2005] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, “Behaviour recognition via sparse spatio-temporal features,” IEEE International Workshop on

Visual Surveillance and Performance Evaluation of Tracking and Surveillance, VS-PETS, pp. 65-72, 2005.

[Doctor 2005] F. Doctor, H. Hagras, and V. Callaghan, "A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments," IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 35, no. 1, pp. 55-65, 2005.

[Doctor 2005] F. Doctor, H. Hagras, and V. Callaghan, "An Intelligent Fuzzy Agent Approach for Realising Ambient Intelligence in Intelligent Inhabited Environments," IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 35, no. 1, pp. 55–65, 2005.

[Doctor 2005] F. Doctor, H. Hagras, and V. Callaghan, "A type-2 fuzzy embedded agent to realise ambient intelligence in ubiquitous computing environments," Journal of Information Sciences, vol. 171, no. 4, pp. 309–334, 2005.

[Dubois 2000] D. Dubois, W. Ostasiewicz, and H. Prade, "Fuzzy sets: history and basic notions," in D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, Kluwer Academic Publishers, Dordrecht, 2000, pp. 80–93.

[Dunn 1973] J. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact, well separated cluster," Cybernetics, vol. 3, no. 3, pp. 32–57, 1973.

[Efros 2003] A. Efros, A. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," IEEE International Conference on Computer Vision, pp. 726-733, 2003.

[Erol 2006] 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.

[Elgammal 2003] A. Elgammal, V. Shet, Y. Yacoob, and L. S. Davis, “Learning dynamics for exemplar-based gesture recognition,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 571-578, 2003.

[Fleet 1994] J. Fleet and S. Beauchemin, “Performance of optical flow techniques,” International Journal of Computer Vision, vol. 12, no. 1, pp. 43–77, 1994.

[Fathi 2008] A. Fathi and G. Mori, “Action recognition by learning mid-level motion features,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008.

[Foroughi 2008] H. Foroughi, BS. Aski, and H. Pourreza, “Intelligent video surveillance for monitoring fall detection of elderly in home environments,” 11th International Conference on Computer and Information Technology, ICCIT, pp. 219-224, 2008.

[Goldberg 1989] D. Goldberg, “Genetic Algorithms in Search, and Machine Learning,” Machine learning, vol. 3, no.2, pp. 95-99, 1989.

[Goldberg 2002] D. Goldberg, “The Design of Competent Genetic Algorithms: Steps toward a Computational Theory of Innovation,” Kluwer Academic Publishers, Dordrecht, 2002.

[Gowsikhaa 2014] D. Gowsikhaa, S. Abirami and R. Baskaran, “Automated human behaviour analysis from surveillance videos: a survey,” Artificial Intelligence Review, vol.42, no.4, pp.747-765, 2014.

[Han 2013] J. Han, L. Shao, D. Xu and J. Shotton, “Enhanced computer vision with microsoft kinect sensor: A review,” IEEE Transactions on Cybernetics, vol. 43, no. 5, pp. 1318-1334, 2013.

[Holland 1975] J. Holland, “Adaptation in Natural and Artificial Systems: an introductory analysis with applications to biology, control, and artificial intelligence” University of Michigan Press, 1975.

[Henricksen 2006] K. Henricksen, J. Indulska, “Developing context-aware pervasive computing applications: Models and approach,” Pervasive and Mobile Computing, vol.2, no.1, pp.37–64, 2006.

[Hagras 2003] H. Hagras, V. Callaghan, M. Cofley, and G. Clarke, “A hierarchical fuzzy-genetic multi-agent architecture for intelligent buildings next term online learning, adaptation and control,” International Journal of Information Sciences, vol. 150, no. 1–2, pp. 33–57, 2003.

[Hagras 2004] H. Hagras, V. Callaghan, M. Colley, G. Clarke, A. Pounds-Cornish and H. Duman, “Creating an ambient-intelligence environment using embedded agents,” IEEE Transaction on Intelligent Transport System, vol. 19, no. 6, pp. 12–20, Nov.–Dec. 2004.

[Hagras 2004] 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.

[Hagras 2007] H. Hagras, “Type-2 FLCs: a new generation of fuzzy controllers,” Computational Intelligence Magazine, IEEE, vol. 2, no. 1, pp. 30-43, 2007.

[Hassanien 2009] A. E. Hassanien, J. H. Abawajy, A. Abraham, and H. Hagra“Pervasive computing: innovations in intelligent multimedia and applications,” Springer Science & Business Media, 2009.

[Hoey 2010] J. Hoey, K. Zutis, V. Leuty amd A. Mihailidis, “A tool to promote prolonged engagement in art therapy: design and development from arts therapist requirements,” Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility, pp. 211-218, 2010.

[Nait-Charif 2004] H Nait-Charif, SJ McKenna, “Activity summarisation and fall detection in a supportive home environment,” Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 4, IEEE, 2004

[Javed 2005] O. Javed, K. Shafique, and M. Shah, “Appearance modeling for tracking in multiple non-overlapping cameras,” Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 26-33. June 2005.

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

[Jiang 2012] Z. Jiang, Z. Lin, L.S. Davis, “A tree-based approach to integrated action localization, recognition and segmentation,” Trends and Topics in Computer Vision, pp.114-127, 2012.

[Ji 2010] X. Ji, H. Liu, Y. Li, “Viewpoint insensitive actions recognition using hidden conditional random fields,” Knowledge-Based and Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, pp.369-378, 2010.

[K. DeJong 1988] K. DeJong, “Learning with genetic algorithms: An overview,” Machine learning, vol. 3, no. 2-3, pp. 121-138, 1988.

[Kaveh 2010] A. Kaveh and S. Talatahari, “A discrete big bang-big crunch algorithm for optimal design of skeletal structures,” Asian Journal of Civil Engineering, vol. 11, no. 1, pp. 103-122, 2010.

[Katz 2003] B. Katz, J. Lin, C. Stauffer, E. Grimson, “Answering questions about moving objects in surveillance videos, ” AAAI Spring Symposium on New Directions in Question Answer, 2003.

[Kleihorst 2008] R. Kleihorst, R. Radke (Eds.), in: Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras, 2008.

[Kleihorst 2007] R. Kleihorst, R. Radke (Eds.), in: Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras, 2007.

[Ke 2007] Y. Ke, R. Sukthankar, and M. Hebert, “Event detection in crowded videos," International Conference on Computer Vision, ICCV, pp. 1-8, 2007.

[Kumbasar 2011] 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.

[Klaser 2008] A. Klaser, M. Marszalek, and C. Schmid, “A spatio-temporal descriptor based on 3d gradients,” The British Machine Vision Conference, pp. 275-1, 2008.

[Klir 2001] G. J. Klir, and B. Yuan, “Fuzzy sets and fuzzy logic theory and applications,” Prentice-Hall of India, New Delhi, 2001.

[Kim 2010] E. Kim, H. Sumi, and C. Diane, “Human activity recognition and pattern discovery,” Pervasive Computing, vol.9, no.1, pp.48-53, 2010.

[Lara 2013] O. Lara, and A. Miguel, “A survey on human activity recognition using wearable sensors,” Communications Surveys and Tutorials, IEEE, vol.15, no.3, pp.1192-1209, 2013.

[Laptev 2003] I. Laptev and T. Lindeberg, “Space-time interest points,” IEEE International Conference on Computer Vision, vol. 64, no. 2-3, pp.107-123, 2003.

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

[Laptev 2007] I. Laptev and P. Perez, “Retrieving actions in movies,” International Conference on Computer Vision, ICCV, pp. 1-8, 2007.

[Laptev 2008] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008.

[Levinson 1997] R. Levinson, “The Planning and Execution Assistant and Trainer (PEAT),” The Journal of head trauma rehabilitation, vol. 12, no. 3–4, pp. 769–775, 1997.

[Liang 2000] Q. Liang and J. Mendel, “Interval type-2 fuzzy logic systems: theory and design”, IEEE Trans. on Fuzzy Systems, vol. 8, pp. 535-550, Oct. 2000.

[Lin 2006] C. Lin, M. Chiu, C. Hsiao, R. Lee, and Y. Tsai, “Wireless health care service system for elderly with dementia,” IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 4, pp. 696–704, 2006.

[Liu 2008] J. Liu and M. Shah, “Learning human actions via information maximization,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008.

[Li 2010] W. Li, Z. Zhang, and Z. Liu, “Action recognition based on a bag of 3d points,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, pp. 9-14, 2010.

[Lee 1990] C. Lee, “Fuzzy Logic in Control Systems: Fuzzy Logic Controller—Part 1”, IEEE transactions on systems, man and cybernetics, Vol. 20, No. 2, pp. 404-418, 1990.

[Mendel 1995] J. Mendel, “Fuzzy logic systems for engineering: a tutorial,” Proceedings of the IEEE, vol. 83, no. 3, pp. 345-377, 1995.

[Mendel 2000] J. Mendel, “Uncertainty, fuzzy logic, and signal processing,” Signal Processing, vol. 80, no. 6, pp.913-933, 2000.

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

[Mendel 2006] J. Mendel, H. Hagras, R. John, “Standard background material about interval type-2 fuzzy logic systems that can be used by all authors,” http://ieee-cis.org/_files/standards.t2.win.pdf, 2006.

[Melin 2010] P. Melin, O. Mendoza, and O. Castillo, “An improved method for edge detection based on interval type-2 fuzzy logic,” Expert Systems with Applications, vol. 37, no. 12, pp.8527-8535, 2010.

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

[Mehrjerdi 2014] Y. Mehrjerdi, “RFID: A Bibliographical Literature Review with Future Research Directions,” International Journal of Industrial Engineering, vol. 25, no. 2, pp.151-190, 2014.

[Miller 1998] J. Miller, M. Smith, M. Mccauley, “Crew fatigue and performance on US coast guard cutters,” no. CG-D-10-99, US Department of Transportation, 1998.

[Maierdan 2013] M. Maierdan, K. Watanabe, and S. Maeyama, “Human behavior recognition system based on 3-dimensional clustering methods,” 13th International Conference on Control, Automation and Systems (ICCAS), pp. 1133-1137, 2013.

[Man 2000] K. Man, K. Tang, S. Kwong, “Genetic Algorithms, Concepts and Designs, Avec disquette” Springer Science & Business Media, pp. 5–10, 2000.

[Nambu 2005] M. Nambu, K. Nakajima, M. Noshiro, and T. Tamura, “An algorithm for the automatic detection of health conditions,” Engineering in Medicine and Biology Magazine, IEEE, vol. 24, no. 4, pp. 38–42, 2005.

[Niebles 2007] J. Niebles and L. Fei-Fei, “A hierarchical model of shape and appearance for human action classification,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[Nowozin 2007] S. Nowozin, G. Bakir, and K. Tsuda, “Discriminative subsequence mining for action classification,” International Conference on Computer Vision, ICCV, pp. 1-8, 2007.

[Ni 2009] B. Ni, S. Yan, and A. Kassim, “Recognizing human group activities with localized causalities,” IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1470-1477, 2009.

[Ni 2013] B. Ni, W. Gang, and M. Pierre, “Rgbd-hudaact: A color-depth video database for human daily activity recognition,” Consumer Depth Cameras for Computer Vision, pp. 193-208, Springer London, 2013.

[Popoola 2012] OP. Popoola, and K Wang, “Video-based abnormal human behavior recognition—A review” IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 42, no. 6, pp. 865-878, 2012.

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

[Pedrycz 1996] W. Pedrycz (Ed.), “Fuzzy Modelling: Paradigms and Practice,” Kluwer Academic Press, Dordrecht, 1996.

[Reyes 2011] M. Reyes, D. Gabriel, and E. Sergio, “Feature weighting in dynamic time warping for gesture recognition in depth data,” Proceedings of IEEE International Conference on Computer Vision Workshops, pp. 1182-1188, 2011.

[Rios 2013] A. Peña-Rios, V. Callaghan, M. Gardner, and M. Alhaddad, “Developing xReality objects for mixed-reality environments,” Intelligent Environments, 2013.

[Rusu 2011] R. Rusu, and C. Steve, “3d is here: Point cloud library (pcl),” Proceedings of IEEE International Conference on Robotics and Automation, pp. 1-4, 2011.

[Ryoo 2009] M. Ryoo, and J. 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.

[Shi 2003] Y. Shi, W. Xie, G. Xu, R. Shi, E. Chen, Y. Mao, and F. Liu, “The smart classroom: Merging technologies for seamless tele-education,” IEEE Pervasive Computing, vol.2, no.2, pp.47–55, 2003.

[Shotton 2013] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M Cook and R. Moore, “Real-time human pose recognition in parts from single depth images,” Communications of the ACM, vol. 56, no. 1, pp. 116-124, 2013.

[Siebel 2006] N. Siebel, S. Maybank, “Fusion of Multiple Tracking Algorithms for Robust People Tracking,” Computer Vision-ECCV 2002, Lecture Notes in Computer Science, Volume 2353/2006, pp. 373-387, 2006.

[Stauffer 1999] C. Stauffer and W. Grimson. “Adaptive background mixture models for real-time tracking,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.

[Schuldt 2004] C. Schuldt, I. Laptev, and B. Caputo, “Recognizing human actions: A local svm approach,” International Conference on Pattern Recognition, ICPR, pp. 32-36, 2004.

[Schindler 2008] K. Schindler and L. V. Gool, “Action snippets: How many frames does human action recognition require?,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008.

[Scovanner 2007] P. Scovanner, S. Ali, M. Shah, “A 3-dimensional SIFT descriptor and its application to action recognition,” Proceedings of the 15th International Conference on Multimedia, pp. 357–360, 2007.

[Sepúlveda 2007] R. Sepúlveda, O. Castillo, P. Melin, A. Rodríguez-Díaz, O. Montiel, “Experimental study of intelligent controllers under uncertainty using type-1, and type-2 fuzzy logic,” Information Sciences, Informatics and Computer Science Intelligent Systems Applications, vol. 177, no. 10, pp. 2023–2048, 2007.

[Song 2000] K. Song, L.H. Sheen, “Heuristic fuzzy-neural network and its application to reactive navigation of a mobile robot,” Fuzzy Sets Systems, vol. 110, no. 3, pp. 331–340, 2000.

[Takagi 1985] T. Takagi, M. Sugeno, “Fuzzy identification of systems and its application to modeling and control,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, no. 1, 1985.

[Tham 2014] JS. Tham, YC. Chang, A. Fauzi, and M. Faizal, “Automatic identification of drinking activities at home using depth data from RGB-D camera,” 2014 International Conference on Control, Automation and Information Sciences (ICCAIS), 2014.

[Trivino 2008] 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.

[Thurau 2008] C. Thurau and V. Hlavac, “Pose primitive based human action recognition in videos or still images,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008.

[Vezzani 2010] R.Vezzani, D. Baltieri, R. Cucchiara, “HMM based action recognition with projection histogram features,” Recognizing Patterns in Signals, Speech, Images and Videos. Springer Berlin Heidelberg, pp. 286-293, 2010.

[Vos 2011] T. Vos, “The Video Surveillance Industry – Market size,” online available at Cameramanager.com

[Wan 2011] J. Wan, C. Byrne, G. O’Hare, and M. O’Grady, “Orange alerts: Lessons from an outdoor case study,” Proceedings of 5th International Conference on Pervasive Computing Technologies for Healthcare, IEEE, pp. 446–451, 2011.

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

[Wu 2006] Y. Wu, T. Yu, “A field model for human detection and tracking, ” IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 28, no. 5, pp. 753–765, 2006.

[Wu 2007] 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.

[Wu 2008] W. Wu, L. Au, B. Jordan, T. Stathopoulos, M. Batalin, W. Kaiser, A. Vahdatpour, M. Sarrafzadeh, M. Fang, and J. Chodosh, “The smartcane system: An assistive device for geriatrics”, Proceedings of 3rd international conference on Body area networks, 2008.

[Wren 1997] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: Real-time tracking of the human body,” EEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), 1997.

[Wang 2007] Y. Wang, P. Sabzmeydani, and G. Mori, “Semi-latent dirichlet allocation: A hierarchical model for human action recognition,” International Conference on Computer Vision, ICCV Workshop on Human Motion, pp. 240-254, 2007.

[Wang 2008] Y. Wang and G. Mori, “Learning a discriminative hidden part model for human action recognition,” Neural Information Processing Systems Foundation, NIPS, pp. 1721-1728, 2008.

[Weinland 2008] D. Weinland and E. Boyer, “Action recognition using exemplar-based embed-ding,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-7, 2008.

[Weinland 2010] D. Weinland, R. Ronfard, and E. Boyer, “A survey of vision-based methods for action representation, segmentation and recognition,” Computer Vision and Image Understanding, vol. 115, no. 2, pp. 224-241, 2010.

[WHO 2014], “10 facts on ageing and the life course,” http://www.who.int/features/factfiles/ageing/en/index.html, [Online], reviewed October 2014.

[Yao 2012] B. Yao, Hani Hagras, Mohammed Ghanbari, Mohammed J. Alhaddad, Daniyal Alghazzawi, “A Type-2 Fuzzy Logic Approach for Detecting Human Related Events in Videos,” UK Workshop on Computational Intelligence, UKCI’12, vol.1, no.1, pp.1-6, 2012.

[Yao 2012] B. Yao, H. Hagras, D. A. Ghazzawi and M. J. Alhaddad, “An Interval Type-2 Fuzzy Logic System for Human Silhouette Extraction in Dynamic Environments,” International Conference on Autonomous and Intelligent Systems, AIS, pp. 126-134, 2012.

[Yao 2014] B. Yao, H. Hagras, M. Alhaddad, D. Alghazzawi, “A fuzzy logic-based system for the automation of human behaviour recognition using machine vision in intelligent environments,” Soft Computing, pp.1–8, 2014.

[Yao 2014] B. Yao, H. Hagras, D. Ghazzawi, and M. Alhaddad, “A Type-2 Fuzzy Logic based system for linguistic summarization of video monitoring in indoor intelligent environments," Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 825-833, 2014.

[Yaouanc 2012] J. Yaouanc, J. Poli, "A fuzzy spatio-temporal-based approach for activity recognition," Advances in Conceptual Modeling, pp. 314-323, 2012.

[Yamato 1992] J. Yamato, J. Ohya, K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 379–385, 1992.

[Yuksel 2012] M. E. Yuksel, and A. Basturk, "Application of Type-2 Fuzzy Logic Filtering to Reduce Noise in Color Images," Computational Intelligence Magazine, IEEE, vol. 7, no. 3, pp.25-35, 2012.

[Yokoi 2006] T.Yokoi, H.Fujiyoshi: "Generating a Time Shrunk Lecture Video by Event Detection", Proceedings of IEEE International Conference on Multimedia and Expo (ICME2005), pp.641-644, 2006.

[Zadeh 1965] L. A. Zadeh. "Fuzzy sets," Information and Control, 1965, vol. 8, pp. 338–353.

[Zadeh 1973] L.A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," IEEE Transactions on Systems, Man, and Cybernetics, vol. 3, no.1, pp. 28–44, 1973.

[Zhou 2001] Q. Zhou and J. Aggarwal. "Tracking and classifying moving objects from video," Proceedings of IEEE Workshop on Performance Evaluation of Tracking and Surveillance. Hawaii, USA, 2001.

[Zhou 2008] 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, pp. 1-7. IEEE, 2008.